

**Prüfung der klinisch-diagnostischen  
Genauigkeit einer neuen  
CAD-basierten Analyseoption  
des multiparametrischen Prostata-MRT  
anhand von 104 durch MR-gestützte  
Biopsie gesicherte Kasuistiken**

**Dissertation  
Zur Erlangung des akademischen Grades  
doctor medicinae (Dr. med.)**

Vorgelegt dem Rat der Medizinischen Fakultät  
der Friedrich-Schiller-Universität Jena

Von Anika Brosch, geb. Thon  
Geboren am 22.07.1991 in Mühlhausen

## **Gutachter**

1. Prof. Dr. med. Ulf Teichgräber, Universitätsklinikum Jena
2. apl. Prof. Dr. med. Ansgar Malich, Südharz Klinikum Nordhausen gGmbH
3. apl. Prof. Dr. med. Joachim Böttcher, SRH Wald-Klinikum Gera

Tag der öffentlichen Verteidigung: 02.09.2019

Meiner Familie in Dankbarkeit gewidmet.

## Inhaltsverzeichnis

Abkürzungsverzeichnis .....	5
1 Zusammenfassung .....	7
2 Einleitung .....	9
2.1 Prostatakarzinom und andere Prostataläsionen .....	9
2.2 Europäische Leitlinien zur Vorsorge, Frühdiagnose und Therapieoptionen des Prostatakarzinoms .....	11
2.3 Multiparametrische Magnetresonanztomographie in der Diagnostik der Prostata .....	13
2.4 Prostate Imaging Reporting and Data System PI-RADS™ V1 und V2 .....	15
2.5 Probengewinnung .....	17
2.6 Gleason Score .....	19
2.7 CAD-Software Watson Elementary™ .....	20
3 Ziele der Arbeit .....	22
4 Publierte Originalarbeit .....	23
5 Diskussion .....	44
6 Schlussfolgerung .....	48
Literatur- und Quellenverzeichnis .....	49
Anhang .....	56
Abbildungsverzeichnis .....	56
Danksagung .....	57
Ehrenwörtliche Erklärung .....	58



## Abkürzungsverzeichnis

AAH	-	Atypical Adenomatous Hyperplasia <u>deutsch:</u> atypische adenomatöse Prostatahyperplasie Adenose
ACR	-	American College of Radiology
ADC	-	Apparent Diffusion Coefficient <u>deutsch:</u> „scheinbarer“ Diffusionskoeffizient
ASAP	-	Atypical Small Acinar Proliferation <u>deutsch:</u> atypische mikroazinäre Proliferation
AUC	-	Area Under The Curve <u>deutsch:</u> Areal unter der Kurve
BPH	-	Benign Prostate Hyperplasia <u>deutsch:</u> benigne Prostatahyperplasie
CAD	-	Computer-Aided Detection <u>deutsch:</u> computerunterstützte Detektion
DCE	-	Dynamic Contrast Enhancement <u>deutsch:</u> dynamische Kontrastmittelbildgebung
DRU	-	Digital-rektale Untersuchung
DWI	-	Diffusion-Weighted Imaging <u>deutsch:</u> Diffusionsgewichtete Bilder
ESUR	-	European Society of Urogenital Radiology
IDC-P	-	Intraductal Carcinoma of Prostate <u>deutsch:</u> Isoliertes Intraduktales Karzinom der Prostata
MAI	-	Malignancy Attention Index
mpMRT	-	multiparametrische Magnetresonanztomografie
MR	-	Magnetic Resonance <u>deutsch:</u> Magnetresonanz
MRSI	-	Magnetic Resonance Spectroscopic Imaging <u>deutsch:</u> Magnetresonanzspektroskopie
PCa	-	Prostate Cancer <u>deutsch:</u> Prostatakrebs
PDT	-	Photodynamic Therapy <u>deutsch:</u> Photodynamische Therapie
PIN	-	Prostatische Intraepitheliale Neoplasie
PI-RADS™	-	Prostate Imaging Reporting and Data System

PSA	-	Prostate Specific Antigen <u>deutsch:</u> prostataspezifisches Antigen
Pt	-	Patient
ROC	-	Receiver Operating Curve <u>deutsch:</u> Grenzwertoptimierungskurve
T2-w	-	T2-weighted Sequence <u>deutsch:</u> T2-Wichtung
TRUS	-	Transrectale Ultrasonography <u>deutsch:</u> transrektale Ultraschalluntersuchung
3T	-	3 Tesla

# 1 Zusammenfassung

EINLEITUNG: Aktuell stellt die quantitative Untersuchung der multiparametrischen Magnetresonanztomographien (mpMRT) mit computer-assistierten Diagnosesystemen (CAD) eine Herausforderung in der Diagnostik von Prostatakrebs (PCa) dar. Nach der Einführung der Therapieoption „active surveillance“ ist die genaue Differenzierung des malignen Potentials von Prostatakarzinomen entscheidend für die weiterführende Therapieempfehlung. In dieser Betrachtung ist neben der Läsionsgröße und dem PSA-Wert auch der Gleason Score von essentieller Bedeutung. Für die präzise Krebsprädiktion und die nicht-invasive Vorhersage des Gleason Scores hat die Firma Watson Elementary™ eine mpMRT – basierte Software entwickelt.

ZIELE DER ARBEIT: Die Genauigkeit und die Effizienz des CAD-Systems von Watson Elementary™ wurde auf Basis der Daten von MRT-gestützt-bioptierten, histologisch gesicherten Prostataläsionen untersucht. Der Schwerpunkt der Untersuchung lag auf der Erfassung der Sensitivität und Spezifität in der Erkennung von malignen Läsionen sowie der Prädiktion des Malignitätsgrades bzw. des Gleason Scores.

METHODEN: Im Hinblick auf diese Zielsetzung wurden retrospektiv 104 MRT-gestützt-bioptierte Läsionen (47 Prostatakarzinome und 57 benigne Läsionen) von insgesamt 79 Patienten im Alter von (Mittelwert  $\pm$  Standardabweichung)  $64,61 \pm 6,64$  Jahren unter Verwendung eines Philips Ingenia 3T MRT-Systems untersucht. Die zur Verfügung stehenden Daten wurden zwischen 10/2013 bis 04/2016 in der radiologischen Abteilung des Südharz Klinikums Nordhausen erhoben. Insgesamt stand eine Kohorte von 122 Patienten zur Verfügung. Von dieser wurden 43 Patienten aufgrund eines inkompatiblen Protokolls aus der Studie ausgeschlossen. Zum ausgewerteten Bildgebungsprotokoll gehörte eine T2-gewichtete Sequenz, eine diffusionsgewichtete Sequenz (DWI) mit b-Werten von 0-100-500-800-1000 und Errechnung des Diffusionskoeffizient (apparent diffusion coefficient, ADC) sowie eine T1-gewichtete Sequenz mit dynamischen Aufnahmen (dynamic contrast enhancement, DCE) nach Gabe von 0,1 mmol/kg Gadoteridol (ProHance™). Bei allen Patienten wurde initial eine diagnostische mpMRT-Untersuchung der Prostata durchgeführt und innerhalb von drei Monaten eine MR-gestützte Stanzbiopsie zur histologischen Sicherung angeschlossen. Bei keinem Patienten wurde vor der Untersuchung eine Chemotherapie durchgeführt, jedoch hatten alle Personen einen

suspekten PSA-Wert beziehungsweise eine suspekte PSA-Wertsteigerung. Auf Grundlage der vorliegenden MRT-Daten (T2-w, DWI und DCE) kalkuliert die Software automatisch eine voxel-basierte Perfusionskurve und eine ADC-Mappe, welche auf Grundlage der DWI-Sequenzen erhoben wird. Durch Konvergenz der T2-w-, ADC- und DCE-Werte wurde ein normierter Index errechnet, der Malignancy Attention Index (MAI). Dieser soll als Vorhersage des Gleason Scores fungieren. Die Bewertung der Läsionen erfolgte retrospektiv mit manuell segmentierten, histologisch gesicherten Läsionen als Goldstandard hinsichtlich der Detektion und der Reabilität sowie Validität der Gleason-Vorhersage.

ERGEBNISSE: Das CAD-System fand 36 der 104 Läsionen, davon 22/47 PCa und 14/57 benigne Prostataläsionen. Dies entspricht einer Sensitivität von 46,80% für PCa mit einer falsch negativen Rate von 0,53/pt. Die Spezifität des CAD-Systems liegt bei 75,43%, mit einem positiv prädiktiven Wert von 61,11%, einem negativ prädiktiven Wert von 63,23% und falsch detektierter Rate von 38,89%. Watson Elementary™ erkannte maligne und benigne Läsionen mit gleicher Wahrscheinlichkeit ( $P$  0.06,  $\chi^2$  test). Die Grenzwertoptimierungskurve (ROC) gibt einen schwachen Gleason-Vorhersagewert für den MAI mit einer Fläche unter der Kurve (AUC)  $< 0,65$  an. Durch Grenzwertoptimierung (MAI cut-off = 0,54) war eine Optimierung der Sensitivität/ Spezifität auf 61,7%/ 68,4% möglich. Weiterhin lässt der MAI keine signifikante Korrelation mit dem Gleason Score ( $P$  0.60, Pearson Korrelation) erkennen.

SCHLUSSFOLGERUNG: Auf Grundlage unserer Ergebnisse eignet sich Watson Elementary™ nicht zur alleinigen PCa-Diagnostik und Vorhersage des Gleason Scores. Es besteht die Empfehlung, das CAD-System als „second read“ mit einer kritischen radiologischen Auswertung zu kombinieren.

## 2 Einleitung

### 2.1 Prostatakarzinom und andere Prostataläsionen

ANATOMIE: Die Prostata wird in vier Zonen unterteilt. Die periphere Zone (Außendrüse) umgibt die zentrale Zone. In ihr finden sich 70% der Karzinome. Die Transitionalzone bzw. periurethrale Mantelzone bildet den Ursprung der Prostatahyperplasie und bildet circa 20% der Malignome aus. Dorsokranial der Urethra befindet sich die zentrale Zone, in welcher circa 10% der bösartigen Neubildungen lokalisiert sind. Die aglanduläre anteriore Stromazone findet sich vor der Transitionalzone (Siehe Abbildung 1).

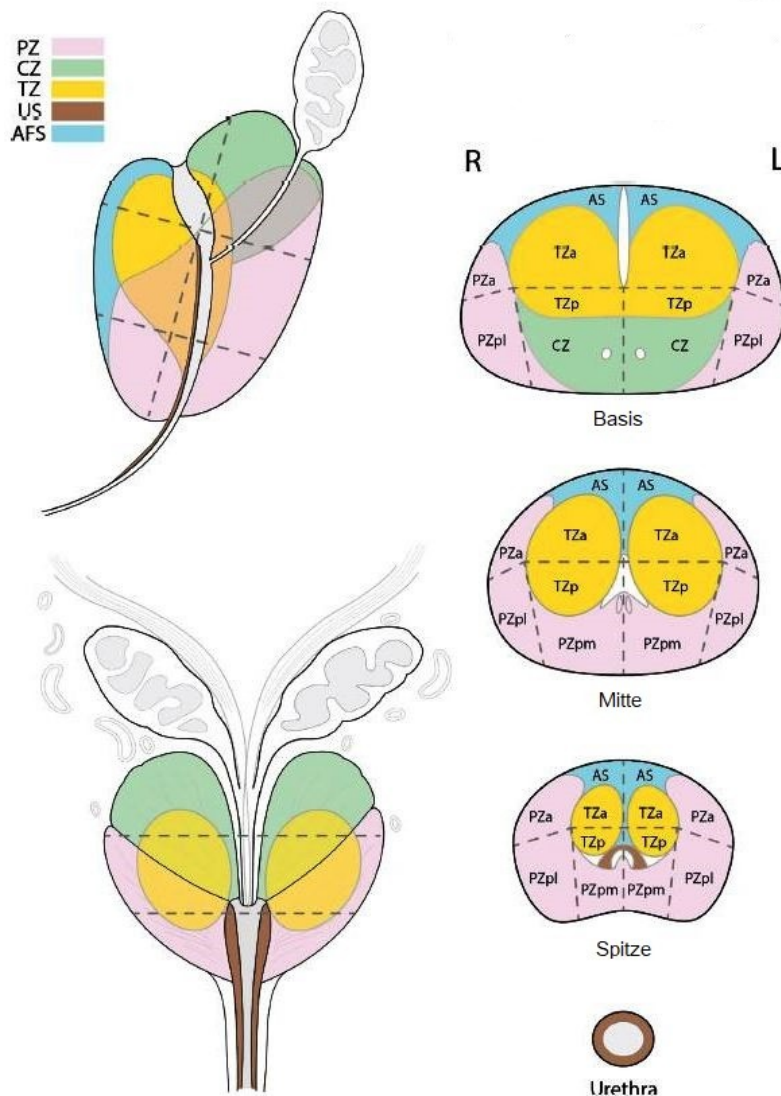


Abb. 1: Aufbau der Prostata und Zoneneinteilung nach PI-RADS V2. PZ= periphere Zone, CZ= zentrale Zone, TZ= Transitionalzone, US= Urethaler Sphincter, AFS= vorderes Fibromuskuläres Stroma, R= rechts, L= Links, (Modifiziert aus Barentsz et al. 2016).

Karzinome sind typischerweise drüsig differenzierte epitheliale Neoplasien mit Expression des Androgenrezeptors und PSA-Positivität. Neben Karzinomen finden sich in der Prostata weitere benigne Läsionen, welche den PSA-Wert beeinflussen können. Bei Männern in der Altersgruppe von 60 bis 69 Jahren leiden 75% unter einer benignen Prostatahyperplasie (BPH), womit diese die häufigste Erkrankung des Mannes im höheren Lebensalter darstellt. Diese führt zudem zu einem Anstieg des Serum PSA-Wertes und maskiert teilweise vorliegende Karzinome. Neben der BPH finden sich in der Prostata weiterhin häufig Entzündungen (Prostatitis), Atrophien und präneoplastische Veränderungen (prostatistische intraepitheliale Neoplasien, PIN/ atypische adenomatöse Prostatahyperplasie Adenose, AAH / atypische mikroazinäre Proliferation, ASAP). Eine PIN zeigt sich vorwiegend in der peripheren Zone und weist eine hohe Korrelation mit dem Adenokarzinom der Prostata auf. Die AAH findet sich überwiegend in der Transitionalzone und wird durch die Basalzellschicht definiert. Liegt eine fragmentierte Basalzellschicht vor, spricht man von einer AAH. Sind keine Basalzellen mehr vorhanden, liegt ein Karzinom bzw. der Übergang zu einem Karzinom vor. Eine ASAP ist eine Mischform zwischen benigner und maligner Formation. In ihr zeigen sich Zellen beider Entitäten, weshalb eine eindeutige Zuordnung schwierig erscheint. Sie stellt den häufigsten Grund für eine Rebiopsie dar (Amann et al. 2016).

EPIDEMIOLOGIE: 2013 wurden laut Robert Koch Institut circa 60.000 Männer mit der Erstdiagnose Prostatakrebs konfrontiert. Insgesamt gab es bereits schätzungsweise 280.000 Männer, mit einer höchstens fünf Jahre zurückliegenden Diagnose (Barnes et al. 2016). Mit 25% ist das Prostatakarzinom die häufigste aller diagnostisierten Krebserkrankungen des deutschen Mannes. Hierbei versterben 11% aller Männer an diesem Tumorleiden. Dies macht das Prostatakarzinom zur zweithäufigsten Krebstodesursache und zur sechst häufigsten Todesursache gesamt. Das mittlere Erkrankungsalter liegt bei 69 Jahren (Leitlinienprogramm 2018).

## **2.2 Europäische Leitlinien zur Vorsorge, Frühdiagnose und Therapieoptionen des Prostatakarzinoms**

SCREENING: Deutsche Männer ab dem 45. Lebensjahr mit einer voraussichtlichen Lebenserwartung von mehr als 10 Jahren haben einen Anspruch auf Früherkennungsmaßnahmen. Sie profitieren von einer frühen Diagnosestellung im organbegrenzten Stadium und einer kurativen Therapie. Zu den Maßnahmen zählt die Anamnese, körperliche Untersuchung und die Digital-rektale Untersuchung (DRU). Die Bestimmung des PSA-Wertes gehört nicht zur Routineuntersuchung. Männer sollten über die Vor- und Nachteile einer PSA-Bestimmung informiert werden. Meist wird ein Ausgangs-PSA-Wert bestimmt, welcher nach einer gewissen Zeit kontrolliert und daraus eine eventuelle Therapie hergeleitet (Leitlinienprogramm 2018).

Der PSA-Wert ist altersabhängig, wird jedoch ein gewisser Schwellenwert überschritten ( $> 4\text{ ng/ml}$ ) oder zeigt sich nach Kontrolle des Wertes eine gewisse Dynamik ( $> 0,35\text{-}0,75\text{ }\mu\text{g/l/Jahr}$ ) wird eine Prostatastanzbiopsie empfohlen (Attenberg et al. 2017, Leitlinienprogramm 2018). Da es bei der Bestimmung des PSA-Wertes jedoch aufgrund unterschiedlicher Faktoren (zum Beispiel Prostatitis, Zystitis oder BPH) vermehrt zu falsch positiven Befunden kommen kann, sind die Aussagekraft und der Stellenwert dieser Untersuchung umstritten (Ilic et al. 2013). Auf der einen Seite steht die effektive, kostengünstige einfache Screening Methode. Auf der anderen Seite darf die hohe falsch-positive Rate der getesteten Männer und die daraus resultierenden weiteren Maßnahmen (Biopsie, Kontrolle des PSA-Wertes) nicht außer Acht gelassen werden. Aus diesen Gründen ist die Notwendigkeit einer genaueren Methode zur Detektion vor allem des frühen Prostatakarzinoms wichtig (Wolf et al. 2010). Eine Prostatabiopsie wird beim Vorliegen eines der nachfolgende Kriterien empfohlen:

- PSA-Wert von  $\geq 4\text{ ng/ml}$  bei erstmaliger Früherkennungsuntersuchung
- suspekte DRU
- auffällige PSA-Dynamik (ohne Wechsel des Bestimmungsverfahrens).

In den aktuellen Leitlinien steht an erster Stelle für die Diagnosesicherung die transrektale Ultraschalluntersuchungs-Biopsie (TRUS-Biopsie). Während dieser transrektalen Ultraschalluntersuchung sollen in der Regel 10-12 Gewebezylinder entnommen werden. Die palpatorisch auffälligen Areale können zusätzlich gezielt biopsiert werden. Bei High-Grade-PIN oder ASAP-Läsionen, isolierten intraduktalen Karzinomen der

Prostata (IDC-P) sowie einem suspekten PSA-Wert bzw. PSA-Verlauf, wird eine erneute Biopsie innerhalb von sechs Monaten empfohlen (Leitlinienprogramm 2018). Diese Gewebeproben stellen die Grundlage zur weiteren Therapieplanung dar.

THERAPIEOPTIONEN: Das Prostatakarzinom hat verschiedene Therapieansätze. Sowohl eine aktive Überwachung (active surveillance), das „Watchful waiting“, als auch eine radikale Prostatektomie, Fokal-, Hormon-, Strahlen- oder Chemotherapie stehen zur Verfügung (Leitlinienprogramm 2018).

Unter „active surveillance“ versteht man eine kurative Methode, durch die der Tumor in den ersten zwei Jahren durch dreimonatliche PSA-Bestimmungen und Digital-rektale Untersuchungen kontrolliert wird. Zeigt sich der PSA-Wert stabil, so kann auf sechsmonatliche Kontrollen umgestellt werden. Die aktuelle Leitlinie 2018 gibt folgende Voraussetzungen für die Anwendung dieser Therapieoption vor:

- PSA-Wert  $\leq 10 \text{ ng/ml}$
- Gleason Score  $\leq 6$
- klinisches Stadium cT1 und cT2 (Wittekind et al. 2005)
- Tumor in  $\leq 2$  Stanzen (bei 10-12 Biopsien)
- $\leq 50\%$  Tumor pro Stanze

Eine erste Rebiopsie wird nach sechs Monaten empfohlen, danach im Abstand von 12-18 Monaten innerhalb der nächsten drei Jahre. Bei stabilem Befund weiterhin alle drei Jahre. Zusätzlich sollten Patienten mit initialem mpMRT und systemischer bzw. gezielter Biopsie eine Rebiopsie mit erneutem mpMRT und systemischer Biopsie nach 12 Monaten erhalten. Für Patienten ohne initiales mpMRT ist ein solches mit systemischer, plus gegebenenfalls gezielter, Biopsie innerhalb von sechs Monaten vorgesehen (Nörenberg et al. 2017, Leitlinienprogramm 2018). Vorteile dieser Therapieoption sind, die Verringerungen von Blutungen und Infektionen durch Vermeidung unnötiger Biopsien und ein Umgehen der damit einhergehenden Heilungsphase (Moore et al. 2013).

„Watchful Waiting“ hingegen beschreibt eine palliative Handhabung. Diese soll die Nebenwirkungen einer Behandlung minimieren, geht jedoch mit einer Verkürzung der Überlebenszeit einher. Hierbei gibt es keine festgelegten Parameter zur Kontrolle. Man kann es jedoch in jedem Stadium der Erkrankung anwenden (Mottet et al. 2016). Die radikale Prostatektomie ergab lediglich bei Patienten unter 65 Jahren und einem mittleren bis hochgradigem Karzinom einen Vorteil im Vergleich zum „Watchful Wai-



ting“ (Heidenreich et al. 2014). Dabei treten nach den Operationen häufig eine erektile Dysfunktion und eine postoperative Inkontinenz auf. Hierbei findet sich nur ein geringer Unterschied zwischen der laparoskopischen und der offenen Prostatektomie (Mottet et al. 2016).

In frühen Erkrankungsstadien wird derzeit als weitere Option die Photodynamische Therapie (PDT) getestet. Diese ist eine minimal-invasive und klinisch in anderen Bereichen bereits anerkannte Methode (Gheewala T et al. 2017). Die PDT zeichnet sich durch drei Hauptkomponenten aus: den Photosensibilisatoren (durch Licht aktivierbare Substanzen), molekularen Sauerstoff und Licht. Nach Verabreichung (intravenös oder lokal) des Sensibilators, reichert sich dieser tumorselektiv an. Danach kann der Tumor mit Licht einer geeigneten Wellenlänge bestrahlt werden. Hierbei werden durch den Sauerstoff und den vorher verabreichten Sensibilator toxische Substanzen produziert, welche den Tumor schädigen (Jocham D et al. 2000). Diese Therapieform kann eine Zwischenstufe zwischen active surveillance und der radikalen Prostatektomie darstellen. Die PDT wird jedoch noch kritisch evaluiert (Ritch C R. 2017).

Eine weitere, jedoch umstrittene Methode ist die Androgendeprivation bei älteren Patienten mit lokal fortgeschrittenem Prostatakarzinom. Diese „chemische Kastration“ geht nicht nur mit einem erhöhten Risiko für Osteoporose, Diabetes mellitus oder dem Abbau der Muskelmasse einher. Sie zeigt darüber hinaus keine Verlängerung der Lebenszeit (Lu-Yao. 2014).

### **2.3 Multiparametrische Magnetresonanztomographie in der Diagnostik der Prostata**

Um invasive Verfahren, wie die Transperitoneale-Ultraschallgesteuerte Biopsie (TRUS) zu optimieren, spielt die mpMRT eine zunehmende Rolle. So wird in den aktuellen Leitlinien ersichtlich, dass sich initial negativ biopsierte Patienten in den mpMRT-gestützten Biopsien in ca. 10% positiv hinsichtlich Prostatakarzinome zeigen. Die beste Detektionsrate wird durch eine Kombination aus der systemischen und der gezielten Biopsie erreicht. Wichtig hierbei ist, dass die mpMRT keinen Teil der Routinediagnostik darstellt. Sie kann die Primärdiagnostik erweitern. Zusätzlich weist ein mpMRT mit PI-RADS < 3 ein geringes Risiko für falsch negative Ergebnisse auf. In diesen Fällen sollte deshalb weiterhin zu oben genannten Therapieoptionen eine systemische Biopsie als Alternative der PSA-Wertkontrollen angeboten werden (Leit-

linienprogramm 2018). Eine groß angelegte Studie aus Großbritannien zeigte eine relevante Reduktion der Primärbioptierate um 27% durch Einsatz einer mpMRT in der Primärdiagnostik (Nörenberg D et al. 2017). Wichtig hierbei ist nicht nur die Erkennung von Läsionen, sondern eine möglichst genaue Differenzierung zwischen benignen und malignen Defekten (Kozlowski et al. 2006). Einen weiteren Schwerpunkt stellt die Diskriminierung unterschiedlicher Malignitätsstufen und somit die weitere Therapieplanung dar. In verschiedenen Studien wurde gezeigt, dass ein signifikanter Unterschied zwischen einer Gleason 6 und 7 Läsion besteht. So werden Gleason 6 Läsionen als niedrigmaligne Läsionen im Gegensatz zu Gleason 7 Läsionen als maligne-hochmaligne Defekte unterschiedlicher Therapieformen zugeführt (Klotz. 2017, Kirkham. 2006).

Desweiteren stellt sie eine wichtige präoperative Voraussetzung dar. Die mpMRT kann nicht nur für eine Unterscheidung zwischen Läsionen genutzt werden. Man kann durch sie auch eine Information über die Größe, Lokalisation und die lokale Ausbreitung (Staging) der Läsionen gewinnen (Wang et al. 2014, Brierley et al. 2017).

SEQUENZEN: T2-w Sequenzen in drei Ebenen (paraaxial, parakoronar und sagittal) sind vor allem für die morphologische Bildgebung der Prostata notwendig. Hier findet man flüssigkeitsreiches Gewebe hell dargestellt. Die T2-w Sequenz erbringt den besten Kontrast zwischen der Transitional- bzw. Zentralzone und der peripheren Zone. Die ESUR-Leitlinie gibt für 3T MRT-Geräte eine Schichtdicke von 3 mm an. Mit ihr kann man eine gute, zonale Unterteilung der Prostata sowie die Beurteilung eines eventuellen extraprostatistischen Wachstums tätigen. Postentzündliche oder -therapeutische Veränderungen sowie Atrophien können die Beurteilung vor allem in der peripheren Zone erschweren (Barentsz et al. 2012). Mit 88-91% ist die T2-w Sequenz zwar sensitiv aber mit nur 61-67% nicht spezifisch für die Erkennung des Prostatakarzinoms. Aus diesem Grund stehen vor allem zur Erhöhung der Spezifität zusätzlich die DWI- sowie die DCE-Sequenzen zur Verfügung (Nörenberg D et al. 2017). Die Diffusionsgewichtete Sequenz (DWI) stellt die extrazelluläre Beweglichkeit der Wasserstoffatome (Brownsche Molekularbewegung) in unterschiedlichen Anregungswinkeln (flip angles) dar. Auf Grundlage dieser Molekularbewegung auf der axialen Ebene (flip angle 90°) wird aus der DWI ein Diffusionskoeffizient, der sogenannte „Apparent Diffusion Coefficient“ (ADC) bestimmt. Der ADC quantifiziert die freie extrazelluläre Wasserbewegung in  $\text{mm}^2/\text{s}$ . Aufgrund der höheren Zelldichte,

zeigen Malignome eine eingeschränkte Wasserdiffusion, welche die freie Brownsche Molekularbewegung einschränkt (hohe DWI-Werte) und den Diffusionskoeffizienten sinken lässt (niedrige ADC-Werte). Eine große Rolle bei der DWI spielt der sogenannte b-Wert. Der b-Wert verbessert den Kontrast zwischen Geweben mit unterschiedlicher Beweglichkeit der Wasserprotonen (Attenberger et al. 2017). Laut ESUR-Leitlinien sollten als b-Werte 0,100 und 800-1000 s/mm<sup>2</sup> genutzt werden. Neuere Studien haben jedoch den Vorteil von b-Werten, von bis zu 1500-2000 s/mm<sup>2</sup> vor allem für unerfahrenere Radiologen, gezeigt (Manenti et al. 2014). Diese Empfehlung wurde allerdings mit einer höheren falsch negativen Rate assoziiert und befindet sich noch in der experimentellen Phase (Ning et al. 2018). Die DCE nutzt die Perfusion von gadoliniumhaltigem Kontrastmittel. Sie besteht aus einer Reihe von schnellen axialen T1-w Sequenzen vor und nach Gabe von Gadolinium. Das maximale zeitliche Auflösungsvermögen sollte hierbei bei 15 s nach Bolusgabe und einer 3ml/s Applikation von Kontrastmittel liegen. Empfohlen wird die weitere Bilderfassung für mindestens fünf Minuten. Hierbei zeigt erkranktes Gewebe typischerweise ein starkes frühes Enhancement und ein sogenanntes „Wash-out“ Phänomen, welches die Abnahme der Kontrastierung nach Bolusgabe bezeichnet. Die Schichtdicke sollte bei 3T Geräten 4 mm betragen (Nörenberg et al. 2017, Barentsz et al. 2012).

## **2.4 Prostate Imaging Reporting and Data System PI-RADS™ V1 und V2**

Im Jahr 2012 wurde von der European Society for UroRadiology (ESUR) und dem American College of Radiology (ACR) erstmals ein System zur Vereinheitlichung von Prostata MRT-Bildern entwickelt das „Prostate Image-Reporting and Data System Version 1“ (PI-RADS™ V1) (Barrett et al. 2015).

In diesem System bekam jede suspekte Läsion und deren Erscheinungsbild einen Punktwert von 1-5. Zusätzlich bildete man einen Gesamtsummenscore, welcher die Wahrscheinlichkeit für das Vorliegen eines klinisch signifikanten Karzinoms angibt.

PI-RADS V1 wurde vor allem in Europa untersucht und durch verschiedene Studien validiert. Es zeigte sich zwar ein deutlicher Nutzen dieser Standardisierung, dennoch gab es viele Uneinigkeiten bezüglich der Beurteilung. Aus diesem Grund konnte 2014 in Zusammenarbeit der ESUR, dem ACR und anderen Institutionen eine überarbeitete Auflage „PI-RADS V2“ herausgegeben werden (Weinreb et al. 2015).

Das Ziel dieser neuen Auflage ist eine Vereinheitlichung der technischen Anforderungen an mpMRTs, sowie eine Vereinfachung und Standardisierung der Untersu-

chung. Ein Hauptbestandteil ist es, eine Vorhersage des Malignitätsrisikos geben zu können. PI-RADS V2 beinhaltet folgende Aspekte:

- DWI ist dominant für Läsionen der peripheren Zone.
- T2-w ist dominant für Läsionen der Transitionalzone.

Für beide gibt es einen separaten 5 Punkte Plan.

- DCE spielt eine untergeordnete Rolle, bekommt einen 2 Punkte Plan (positiv oder negativ).
- Bei der DWI-Sequenz sind hohe b-Werte notwendig. Aus ihr wird die ADC-Mappe berechnet.
- Die Prostata wird in 39 Sektoren eingeteilt (siehe Abbildung 1).
- Magnetic resonance spectroscopic imaging (MRSI) wird nicht benötigt.
- Die Größe der Läsion wird zur Punkteermittlung beurteilt.

Die Zusammenschau der oben genannten Merkmale führt weiterhin zu einem Punktwert auf der Likert-Skala von eins bis fünf, welcher das potentielle Malignitätsrisiko vorhersagt (siehe Abbildung 2):

- 1: Sehr niedrig (das Vorliegen eines Karzinoms ist sehr unwahrscheinlich)
- 2: Gering (das Vorliegen eines Karzinoms ist unwahrscheinlich)
- 3: Intermediär (das Vorliegen eines Karzinoms ist fragwürdig)
- 4: Hoch (das Vorliegen eines Karzinoms ist wahrscheinlich)
- 5: Sehr hoch (das Vorliegen eines Karzinoms ist sehr wahrscheinlich)

### Transitionalzone

T2W			PIRADS Score
1	Homogene Signalintensität		1
2	Umschriebene Hypodensität oder heterogene intraprostatrische Knoten (BPH)		2
3	Heterogene Signalintensität mit unscharfen Rändern, Exklusive 2, 4 oder 5	DWI Score ≤ 4 → 5 →	3
4	Entrundet oder randlos, homogen, moderat hypointens, < 1,5cm		4
5	Siehe 4, > 1,5cm oder extraprostatrische Ausdehnung, invasives Wachstum		5

### Periphere Zone

DWI			PIRADS Score
1	Keine suspekten Areale in ADC oder DWI gesehen		1
2	Unklare Hypodensität in ADC		2
3	Fokale milde/moderate Hypodensität in ADC und isointense/milde Hyperdensität DWI	DCE NEGATIV → POSITIV →	3
4	Fokal demarkierte Hypodensität in ADC und Hyperdensität in DWI, < 1,5cm		4
5	Siehe 4, > 1,5cm oder extraprostatrische Ausdehnung/ invasives Wachstum		5

Abb. 2: Übersicht über die Zusammensetzungen der Bewertungen in PI-RADS V2 der unterschiedlichen Zonen und Sequenzen (Modifiziert aus Barentsz et al. 2016).

## 2.5 Probengewinnung

Um oben genannte Proben zu gewinnen, stehen zwei verschiedene Möglichkeiten zur Verfügung: die gezielte und die zielgerichtete Biopsie. Zudem können beide Varianten in Kombination eingesetzt werden.

1. Mittels MRT-gestützter Prostatabiopsie, werden aus einer definierten Läsion gezielt 1-3 Stanzproben entnommen. Hierbei wird der Patient unter laufender MRT-Kontrolle gezielt punktiert. Zusätzlich kann dazu eine CAD-Software zur Fusionierung und Kalkulation der Koordinaten, der vorher detektieren Läsionen genutzt werden. Mit Hilfe dieser Kalkulationen wird eine Führungshülse für die Punktionsnadel positioniert (siehe Abbildung 3).

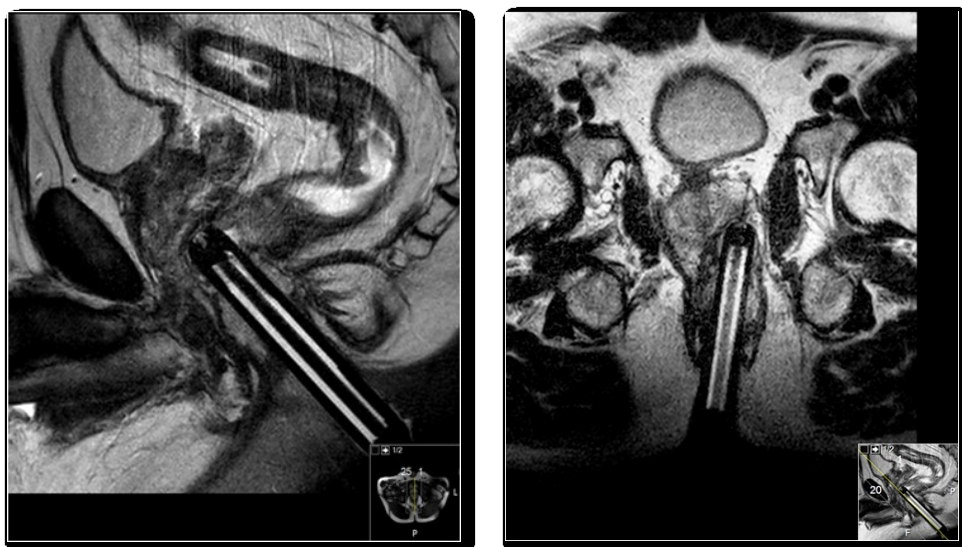


Abb. 3: Prostatastanzbiopsie. Links: T2-w sagittal und parakoronal mit Darstellung der Führungshülse im Enddarm und der angezielten Prostataläsion. Rechts: Dokumentation der Biopsiestelle (in der linken peripheren Zone) mit ausgefahrener Nadel (Malich, nicht publiziert).

2. Die Fächerbiopsie zeichnet sich dadurch aus, dass die Prostata nach einem bestimmten Muster bioptiert wird. Die Leitlinie der DGU empfiehlt dabei, dass 10-12 Gewebezylinder entnommen werden sollten. Bei palpatorisch auffälligen Befunden, können diese zusätzlich dazu punktiert werden. Hierbei können Karzinome oftmals nicht detektiert werden (Börgermann et al. 2014).
3. Eine Kombination aus beiden Varianten ist die MRT/TRUS-Fusionsbiopsie. Bei dieser werden die mpMRT-Bilder mit TRUS-Bildern fusioniert und damit eine gezielte 12-fach Biopsie aus den suspekten Läsionen gewonnen (Brock et al. 2015).

Erstere hat zwei Vorteile: Sie liefert eine bessere Aussage zur Operabilität des Befundes bzw. zum Kapseleinbruch und per continuitatem Infiltration des perinealen Fettkörpers, der Urethra oder des Rektums sowie eine genauere Lokalisation der Läsion. Dies ermöglicht eine eindeutigere operative Entscheidung zum Erhalt des neurovaskulären Bündels (Nörenberg et al. 2017).

Ein entscheidender Unterschied zwischen Variante eins und zwei liegt darin, dass die gezielte Punktion kein „blindes Verfahren“ ist. Weitere Vorteile gegenüber der Fächerbiopsie sind zudem weniger Komplikationen (Blutungen, Infektionen) und eine erhöhte Treffsicherheit, sowie die Reduktion der falsch negativen Biopsien. Es folgt eine bessere Risikoklassifizierung und eine Verringerung von Rebiopsien (Bjurlin et al. 2014). Mit einer Sensitivität von nur 53% weist die Fächerbiopsie eine hohe falsch negative Rate auf, das heißt die erfolglose Detektion von existierenden malignen Läsionen.

sionen (Brock et al. 2015). Als eine mögliche Ursache hierfür wird die nahe an der Harnblase liegende Prostatazone (anterior und apikal) genannt. Aufgrund der höheren Wahrscheinlichkeit einer Urethraverletzung werden diese Lokalisationen häufig nicht erfasst. Weiterhin kann nur eine bedingte Einschätzung der Ausdehnung der Läsion getätigt werden. Hinzu kommt oftmals die Unterschätzung des Gleason Scores (Nörenberg et al. 2017). Aus diesem Grund müssen circa ein Drittel der Männer innerhalb von fünf Jahren erneut biopsiert werden, in der wiederholten Biopsie werden dann wiederum 13-41% maligne Diagnosen gestellt (Brock et al. 2015).

## 2.6 Gleason Score

Speziell für die Prostata wurde ein Gradingssystem entwickelt, der sogenannte Gleason Score. Dieser dient zur Beurteilung der verschiedenen histologischen Wachstumsmuster des Prostatakarzinoms. Besonders hierbei ist, dass das Scoringssystem sowohl das vorherrschende als auch das zweitrangige Wachstumsmuster berücksichtigt. „Die Grundmuster werden nach abnehmendem Grad der histologischen Differenzierung mit steigender Punktzahl (1 bis 5) belegt, Der niedrigste Malignitätsgrad ist durch jeweils einen Punkt des primären (vorherrschenden) und sekundären (zweithäufigsten) Grundmusters charakterisiert ( $1 + 1 = 2$ )“ (Amann et al. 2016) (siehe Abbildung 4).

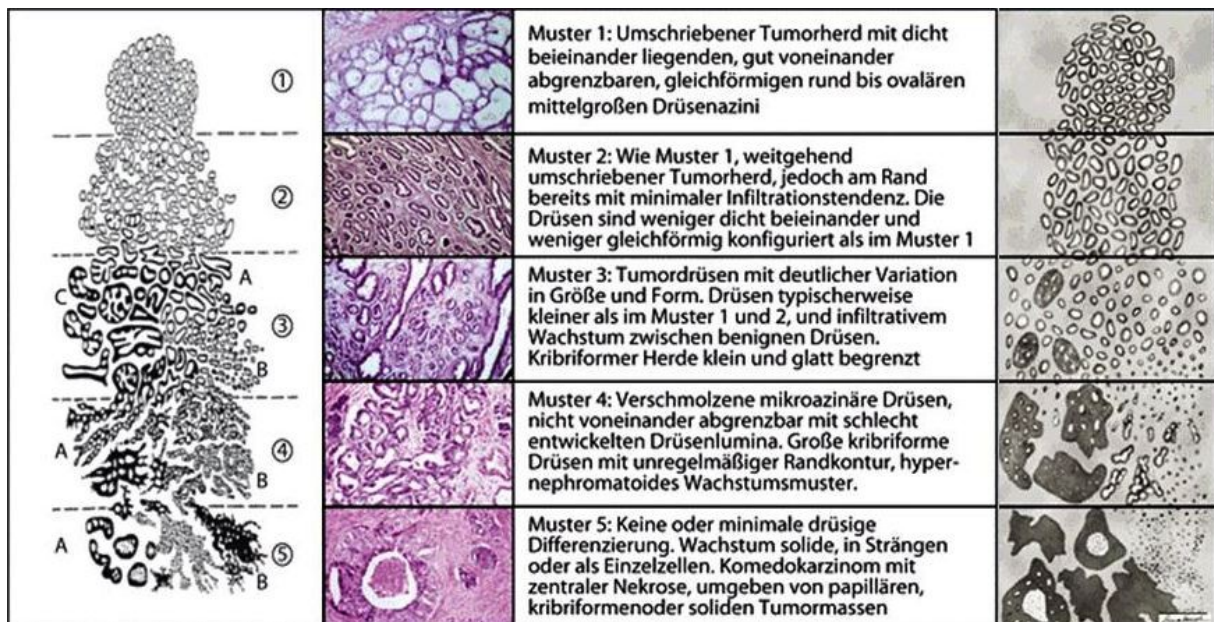


Abb. 4: Gleason Score. Links: Schematische Darstellung Rechts: Histologische Darstellung mit Beschreibung der einzelnen Muster (Modifiziert aus Börgermann et al. 2014).

Somit liegt der höchste Gleason Score bei 10 (5 + 5). Jedes Muster hat eine eigene, genaue Definition. Ein Gleason Score von 6 und 7a (3 + 4) wird einem mittleren Risiko zugeordnet, wohingegen Werte ab 7b (4 + 3) – 10 zu den High-risk Karzinomen gezählt werden. Dieser Wert spielt in der Therapieplanung eine wichtige Rolle. So können Karzinome mit einem Gleason Score bis 7a und weiteren günstigen Faktoren (siehe Punkt 2.2 Therapieoptionen) mit active surveillance beobachtet werden, wohingegen aggressive Tumoren in der Regel radikal operiert werden, in seltenen Fällen aber auch andere Therapiekonzepte möglich sind.

## **2.7 CAD-Software Watson Elementary™**

Die untersuchte CAD-Software ist eine statische Anwendung. Hierunter versteht man ein nicht trainierbares System mit festen Parametern. Die mpMRT Bilder werden vollautomatisiert ohne mögliche Intervention des Anwenders in drei Schritten bearbeitet (Roethke et al. 2016):

1. Segmentierung und Fusionierung der Prostata aus den Input-mpMRT Bildsequenzen (T2-w, DWI, DCE).
2. Quantitative Bildbearbeitung Generierung einer ADC-Mappe aus den DWI Sequenzen, Kinetikauswertung aus der DCE-Serie mit Implementation des Tofts-Models (Ktrans, Ve and Kep), Kontrastbearbeitung der Bilder zur Differenzierung unterschiedlicher Gewebebeschaffenheiten.
3. Detektion der (laut interner Systeminformation) statistisch suspekten Areale und Zusammenführung der aktuellen Berechnungen zur Darstellung in einem Datendiagramm.

Am Ende dieser Schritte präsentiert die CAD-Software einen auf Pixeln basierenden Malignitätsvorhersagewert zwischen 0 und 1. Dieser „Malignancy Attention Index“ (MAI) gibt eine Korrelation zum potentiellen Gleason Wert an (Roethke et al. 2016). Zur Visualisierung der suspekten Areale generiert das Programm eine Wärmekarte. In dieser werden suspekte Areale mit warmen Farben (rot) gekennzeichnet, wohingegen benigne Läsionen in kalten Farben (blau) repräsentiert werden. Diese Farbgenerierung wird auf die T2-w MRT Sequenz projiziert (Thon et al. 2017) (siehe Abbildung 5).



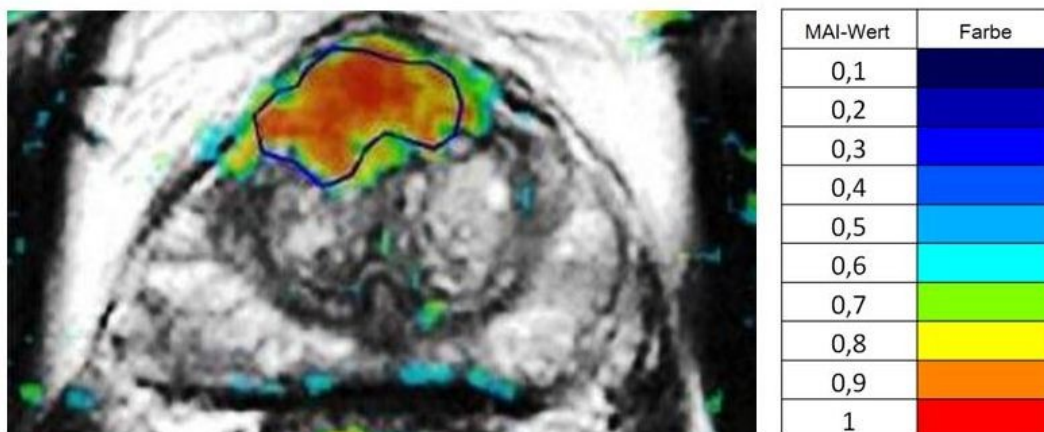


Abb. 5: Malignancy Attention Index. Links: T2-w Sequenz mit generierter MAI Farbskalierung (Thon nicht veröffentlicht), rechts: Zuordnung der einzelnen MAI-Werte zu einer Farbe. (modifiziert aus Roethke et al. 2016)

Wie in Abbildung 6 dargestellt, werden weiterhin die suspekten Areale automatisch in ein Säulendiagramm generiert (siehe Abbildung 6).

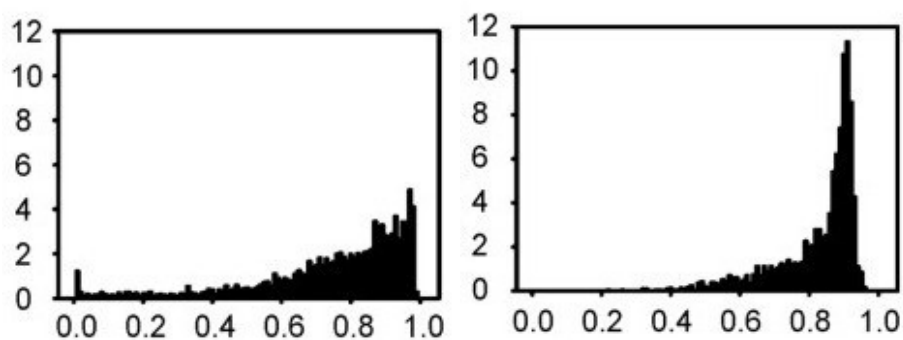


Abb. 6: MAI-Säulendiagramme. x-Achse: MAI-Wert, y-Achse: Vorkommen (%). Links: PCa Gleason 6 Läsion, rechts: PCa Gleason 9 Läsion (Thon et al. 2017).

### **3 Ziele der Arbeit**

Die vorliegende Dissertationsarbeit analysiert das diagnostische Potential des CAD-Systems (Watson Elementary™, Watson Medical, Den Ham, Niederlande) in der Detektion von malignen Prostataläsionen auf der Grundlage von mpMRT-Bildern. Hierfür wurden unterschiedliche Sachverhalte geprüft:

- In wie vielen Fällen kann die Software die arterielle Verlaufskurve richtig definieren?
- Wie hoch ist die Sensitivität des CAD-Systems?
- Wie hoch ist die Spezifität des CAD-Systems?
- Kann der berechnete MAI als Vorhersagewert des Gleason Scores genutzt werden?
- Gibt es eine Korrelation zwischen Detektionsrate und Größe/Volumen der Läsion?

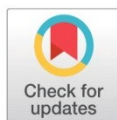
## RESEARCH ARTICLE

# Computer aided detection in prostate cancer diagnostics: A promising alternative to biopsy? A retrospective study from 104 lesions with histological ground truth

Anika Thon<sup>1,2</sup>, Ulf Teichgräber<sup>1</sup>, Cornelia Tennstedt-Schenk<sup>3</sup>, Stathis Hadjimetriou<sup>4</sup>, Sven Winzler<sup>2</sup>, Ansgar Malich<sup>2</sup>, Ismini Papageorgiou<sup>2\*</sup>

**1** Institute of Diagnostic and Interventional Radiology, Department of Experimental Radiology, Jena University Hospital, Friedrich-Schiller University, Jena, Germany, **2** Institute of Radiology, Suedharz Hospital Nordhausen gGmbH, Nordhausen, Germany, **3** Institute for Pathology, Mühlhausen-Pfafferode, Germany, **4** Department of Electrical Engineering and Informatics, Cyprus University of Technology, Limassol, Cyprus

\* ismini.papageorgiou@shk-ndh.de



## Abstract

## OPEN ACCESS

**Citation:** Thon A, Teichgräber U, Tennstedt-Schenk C, Hadjimetriou S, Winzler S, Malich A, et al. (2017) Computer aided detection in prostate cancer diagnostics: A promising alternative to biopsy? A retrospective study from 104 lesions with histological ground truth. PLoS ONE 12(10): e0185995. <https://doi.org/10.1371/journal.pone.0185995>

**Editor:** Tone Frost Bathen, Norges teknisk-naturvitenskapelige universitet, NORWAY

**Received:** March 7, 2017

**Accepted:** September 22, 2017

**Published:** October 12, 2017

**Copyright:** © 2017 Thon et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** The authors received no specific funding for this work.

**Competing interests:** The authors have declared that no competing interests exist.

## Background

Prostate cancer (PCa) diagnosis by means of multiparametric magnetic resonance imaging (mpMRI) is a current challenge for the development of computer-aided detection (CAD) tools. An innovative CAD-software (Watson Elementary™) was proposed to achieve high sensitivity and specificity, as well as to allege a correlate to Gleason grade.

## Aim/Objective

To assess the performance of Watson Elementary™ in automated PCa diagnosis in our hospital's database of MRI-guided prostate biopsies.

## Methods

The evaluation was retrospective for 104 lesions (47 PCa, 57 benign) from 79, 64.61±6.64 year old patients using 3T T2-weighted imaging, Apparent Diffusion Coefficient (ADC) maps and dynamic contrast enhancement series. Watson Elementary™ utilizes signal intensity, diffusion properties and kinetic profile to compute a proportional Gleason grade predictor, termed Malignancy Attention Index (MAI). The analysis focused on (i) the CAD sensitivity and specificity to classify suspect lesions and (ii) the MAI correlation with the histopathological ground truth.

## Results

The software revealed a sensitivity of 46.80% for PCa classification. The specificity for PCa was found to be 75.43% with a positive predictive value of 61.11%, a negative predictive value of 63.23% and a false discovery rate of 38.89%. CAD classified PCa and benign lesions with equal probability ( $P0.06$ ,  $\chi^2$  test).

Accordingly, receiver operating characteristic analysis suggests a poor predictive value for MAI with an area under curve of 0.65 ( $P=0.02$ ), which is not superior to the performance of board certified observers. Moreover, MAI revealed no significant correlation with Gleason grade ( $P=0.60$ , Pearson's correlation).

## Conclusion

The tested CAD software for mpMRI analysis was a weak PCa biomarker in this dataset. Targeted prostate biopsy and histology remains the gold standard for prostate cancer diagnosis.

## Introduction

Prostate cancer (PCa) is the third most common cancer in the total population, representing approximately 11% of all cancer diagnoses [1]. 55–60% of prostate cancer patients are men over 65 years of age. The disease has an excellent survival rate of 94% in the 1<sup>st</sup> year, 85% in 5 years and 84% in 10 years [2]. The relative survival rate reaches 100% in the first 5 years and 95% in 15 years [1], provided that cancer will be diagnosed at the local (stage I, II) or the regional stage (stage III), as occurs with more than 80% of the cases [1]. A 5-year survival rate at stage IV (distant disease) of 28–30% renders the importance of early stage clinical diagnosis indisputable [1,2].

In spite of being faced with criticism [3–6], the prostate-specific antigen (PSA) assay [7,8] has been, and continues to be the most popular and widely applied PCa screening method in practice for the last 30 years. However, invasive methods such as prostate biopsy still prevail as the gold standard for preoperative evaluation, risk-assessment and decision-making between active surveillance, new evolving tissue-preserving strategies and more radical approaches for aggressive disease such as whole-gland radiation, chemotherapy and radical prostatectomy [7–12]. Due to the low PSA specificity and a large number of false positives, there is an increasing need for a non invasive imaging PCa biomarker [9,10].

A promising list of laboratory non invasive PCa-biomarkers is currently under validation but none of them are a part of the clinical routine [5,9,10]. Prostate magnetic resonance imaging (MRI) is, on the other hand, a promising biomarker already in clinical routine for diagnosis and preoperative evaluation of prostatic lesions in patients with elevated PSA [13–17]. MRI is superior to other diagnostic methods for providing information about the size, localization and spread of the disease. Furthermore, the available multiparametric MRI-sequence battery (mpMRI), including high-resolution T2-weighted imaging (T2w), diffusion-weighted imaging (DWI) with apparent diffusion coefficient (ADC) maps and dynamic contrast-enhanced (DCE) sequences, allows for precise localization and detailed anatomical and functional description of prostatic lesions and their neighboring structures [13,18–21]. The non invasive character and the high negative predictive value of mpMRI combined with minimal invasive MRI-guided biopsy [22], justify them as attractive diagnostic tools that obviate more invasive diagnostic methods such as the multicore (not lesion-targeted or systematic) mapping biopsy [10].

Although the sensitivity of mpMRI is constantly increasing, especially in line with technical advances in diffusion imaging [16], it does not yet qualify for the accurate differential diagnosis of PCa from indolent alterations with similar MRI features, such as benign prostate hypertrophy [14]. Moreover, the interpretation of mpMRI requires highly qualified, board-certified



staff, the supply of which is disproportionately low compared to the increased diagnostic demand of the second most common cancer in males. An increasing body of evidence supports the role of automated mpMRI analysis in the form of Computer-Aided Detection (CAD) methods. CAD systems approach MRI modalities quantitatively and allow for information convergence into statistical pipelines that are adjusted to predict malignancy [13].

A recently commercialized, automated analysis tool for the assessment of prostate cancer in mpMRI (Watson Elementary™, Watson Medical, Den Ham, The Netherlands) has achieved high sensitivity and specificity in its first evaluation [23]. The Watson Elementary™ method is tuned up to predict the malignancy grade with an mpMRI-based Gleason score correlate, termed Malignancy Attention Index (MAI). In this study, Watson Elementary™ was retrospectively assessed in our hospital's database of 104 prostate lesions with histological ground truth after MRI-guided biopsy, and showed a low sensitivity for PCa detection, which was not superior to the observational diagnosis. Our results are compared with previous related studies.

## Materials and methods

### Ethics

All patient data were derived from the prostate database of the Suedharz Hospital Nordhausen. Data were analyzed retrospectively, fully anonymized, in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its amendments as well as with the guidelines of the Ethical Committee for clinical studies of the University of Jena. Due to the retrospective character of the study, the ethical committee has waived the mandate for obtaining a legally effective informed consent from the included subjects. Accordingly, therapeutic decisions were not influenced by the outcome of this study.

### Study design

The evaluation was retrospective for 104 histologically characterized biopsy cores obtained with MRI-guided prostate biopsy (47 malignant, 57 benign) from 79 patients aged 48–80 years (average  $\pm$   $\sigma$  64.62  $\pm$  6.64 y.o.), 35 with malignant and 44 with benign lesions, scanned in our department between 10/2013 and 4/2016 (Table 1, Fig 1). All patients were examined with suspicion for prostate cancer based on elevated PSA assay after a negative systematic biopsy and none of them had previously received chemotherapy for prostate cancer treatment. From a total of 122 patients, 43 patients were excluded from this study due to protocol mismatch with the software's technical requirements: in 27 patients, rejection was due to static field strength inconsistencies (1.5T excluded); in one patient, due to anatomic malposition of the prostate; in eight patients, the arterial reference curve in the DCE sequences was insufficient; three patients were rejected due to motion artifacts; in two cases, the fusion step failed (see later); and in one case the arterial input curve could not be defined. Images from a single patient were not accepted by the software for an unknown reason. No patient- or lesion-related inclusion/exclusion criteria applied.

All lesions were graded by 2 radiologists; one with intermediate experience, and one board-certified radiologist, according to the Prostate Imaging Reporting And Data System (PI-RADS™) v1 and -v2. MRI-guided transrectal needle biopsies were always performed in less than 3 months post diagnosis. Histological characterization and Gleason grading on H&E stained sections followed.

Watson Elementary™ (installed on 11.5.2016 by Watson Medical, Den Ham, The Netherlands) was tested with a manual generation of regions of interest (ROI) in diagnostic MRI series encompassing targeted biopsy cores. The diagnostic accuracy hypothesis tested the ability of CAD to detect known, manually drawn lesions. Therefore all lesions without histological

Table 1. Patient demographics and database description.

	Age (y)	N	n	Classified (%; n)	PIRADS V01 (%; n)			PIRADS V02 (%; n)				Histology
					3	4	5	2	3	4	5	
<b>Total</b>	64,61±6,64	79	104	35; 36								
<b>Malignant</b>	61,64± 0	1	2	0; 0	0; 0	50; 1	50; 1	0; 0	0; 0	50; 1	50; 1	Gleason 5
	65,70± 6,01	6	7	14; 3	0; 0	43; 3	57; 4	0; 0	0; 0	43; 3	57; 4	Gleason 6
	64,67± 5,12	18	25	55; 12	0; 0	40; 10	60; 15	0; 0	0; 0	64; 16	36; 9	Gleason 7
	65,97± 10,57	5	6	14; 3	0; 0	17; 1	83; 5	0; 0	0; 0	67; 4	33; 2	Gleason 8
	69,06± 5,32	4	5	18; 4	0; 0	0; 0	100; 5	0; 0	0; 0	0; 0	100; 5	Gleason 9
	62,54±0	1	2	0; 0	0; 0	50; 1	50; 1	0; 0	0; 0	100; 2	0; 0	Gleason 10
<b>Total malignant</b>	65,38±6,18	35	47	61; 22	0; 0	34; 16	66; 31	0; 0	0; 0	55; 26	45; 21	
<b>Benign</b>	63,79± 4,77	10	15	21; 3	0; 0	73; 11	27; 4	7; 1	7; 1	60; 9	27; 4	Prostatitis
	62,78± 8,75	15	20	36; 5	15; 3	55; 11	30; 6	0; 0	10; 2	80; 16	10; 2	BPH
	67,52± 6,23	4	6	14; 2	0; 0	50; 3	50; 3	0; 0	0; 0	83; 5	17; 1	Prostatic tissue
	65,99± 8,90	2	2	7; 1	0; 0	100; 2	0; 0	0; 0	50; 1	50; 1	0; 0	Atrophy
	63,61± 0	1	1	0; 0	0; 0	100; 1	0; 0	0; 0	0; 0	100; 1	0; 0	FMD
	64,25± 7,40	12	13	21; 3	0; 0	54; 7	46; 6	8; 1	8; 1	70; 9	15; 2	ASAP
<b>Total benign</b>	64,01±6,73	44	57	39; 14	5; 3	61; 35	33; 19	4; 2	9; 5	72; 41	16; 9	

Software evaluation was based on a series of 104 histologically confirmed prostatic lesions from 79 patients. Multiparametric MRI evaluation was performed in diagnostic series acquired within 3 months before the biopsy. Lesions have been scored according to the Prostate Imaging Reporting And Data System (PI-RADS™) v1 and -v2. All values are reported in mean,  $\sigma$ . N, patients; n, lesions; BPH, Benign Prostate Hyperplasia; FMD, Fibromuscular Dysplasia; ASAP, Atypical Small Acinar Proliferation.

<https://doi.org/10.1371/journal.pone.0185995.t001>

ground truth, including new lesions identified by the software, were neglected and did not influence the statistics of the current study.

Evaluation of the diagnostic accuracy followed with estimation of the sensitivity, specificity, Positive Predictive Value (PPV) and False Negative Rate (FNR). Implementation of Receiver Operatic Characteristic (ROC) analysis and estimation of the Area Under the Curve (AUC) allowed for the definition of optimal cut-off diagnostic values using Youden statistics (see also *Statistics*).

## MRI-guided prostatic biopsy

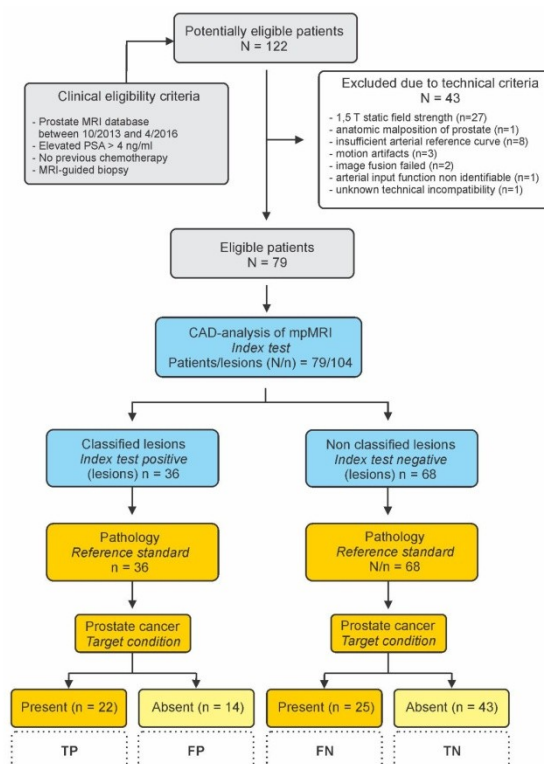
Malignancy suspect lesions according to PI-RADS™ scoring were transrectally biopsied under stereotactic MRI-guidance with a Philips Ingenia 3.0T MR system using a dStream Torso body coil with Flex Coverage anterior and posterior coils that allow for a 32-channel, 60 cm body coverage (Philips North America Corporation, Andover, MA USA).

Briefly, a stereotactic needle-frame with x-y-z freedom was fixed onto the patient table. Target lesions were (re-)allocated and the sampling position was planned with the frame-dedicated software (DynaCAD, Invivo Corporation, Gainesville, FL, USA). The accurate position of the biopsy needle-tip was confirmed with a T2w TSE HR sequence.

## Image acquisition for mpMRI

Both biopsies and diagnostic imaging were performed on the same Philips Ingenia 3.0T MR-system within a time interval of 3 months. The following axially orientated image set was used for CAD analysis (Table 2):

1. T2 weighted high-resolution Turbo Spin Echo imaging (T2 TSE HR) (Fig 2Ai, 2Bi and 2Ci)



**Fig 1. Flow of participants through the study.** STARD diagram to demonstrate the flow of participants through the study. Downstream of the "Index test" the analysis is lesion-based due to detection of multiple lesions in some patients. PSA, Prostate Specific Antigen; CAD, Computer-Aided Detection; TP, True Positive; TN, True Negative; FP, False Positive; FN, False Negative.

<https://doi.org/10.1371/journal.pone.0185995.g001>

- Diffusion weighted Echo Planar Imaging (DWI EPI) at 5 different b-values (b0-100-500-800-1000)
- T1 weighted Fast Field Echo (T1-FFE) with dynamic contrast enhancement (DCE) in 25 repetitions with 13.35 s temporal resolution and 7 s delay, which corresponded to at least two baseline (zero contrast) points. A body weight-adjusted bolus of gadoteridol 0.1 mmol/kg (ProHance<sup>®</sup>, Bracco S.p.A., Milan, Italy) was injected at 3 ml/s flow rate. Watson Elementary<sup>™</sup> provides only visual access to the DCE curve in.pdf format, which is not indicated for publication purposes. Therefore, for demonstration purposes, example DCE curves from selected lesions were manually exported (Fig 2Aiii and 2Aiv, 2Biii and 2Biv, 2Ciii and 2Civ).

### Description of Watson Elementary<sup>™</sup>

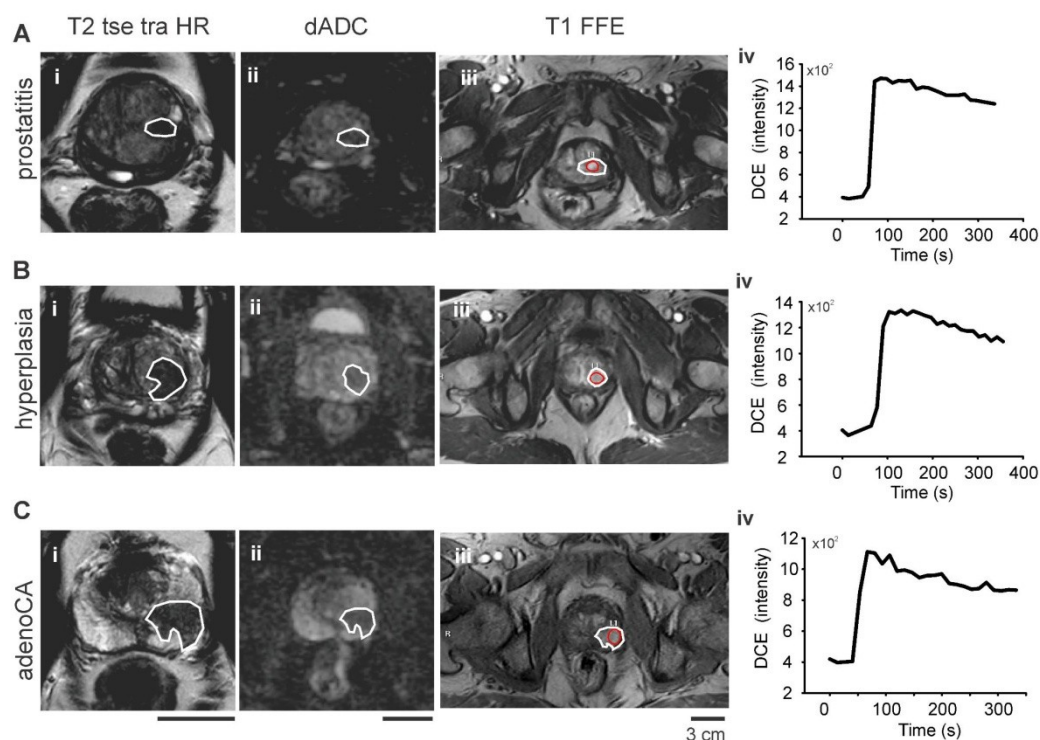
Watson Elementary<sup>™</sup> software implements a fully automated 3-step-method, previously reported in detail [23]. Below is a brief description of its image processing steps:

**Table 2. Sequences used for multiparametric magnetic resonance imaging.**

Parameter/Sequence	T2 TSE HR	DWI EPI 20	T1 FFE with DCE
TR(ms)/TE(ms)	4811/120	4227/63	5.5/1.93
Flip angle (°)	90	90	15
EPI factor		single shot technique	
Number of averages	1	4	1
b values s/mm <sup>2</sup>		0-100-500-800-1000	
Slice thickness (mm)	3	3	3
FOV AP×FH×RL (mm)	260×75×160	256×75×256	262×120×262
Voxel size AP×RL×FH (mm <sup>3</sup> )	0.65×0.65×3	1.8×1.8×3	1.02×1.41×3
Aquisition time (min:s:ms)	3:12	5:25	0:13:35 pro sequence, 25 dynamic sequences 5:31 in total
Enhancer	Gadoteridol (ProHance®) 0.1 mmol (0.2 ml)/kg, injection rate 3 ml/s		

Imaging was performed at 3T field strength. TR, repetition time; TE, echo time; EPI, Echo-Planar Imaging factor and number of k-spaces collected in a single shot; FOV, Field Of View; AP, Anterior to Posterior; FH, Foot to Head; RL, Right to Left.

<https://doi.org/10.1371/journal.pone.0185995.t002>



**Fig 2. Sequences used for multiparametric MRI (mpMRI) imaging according to PI-RADS™.** Sample images from lesions (white ROIs in A-C) histologically characterized as (A) prostatitis (B) hyperplasia (BPH) and (C) adenocarcinoma (PCA). (i) T2w TSE images, (ii) apparent diffusion coefficient (ADC) tables based on DWI b0-100-500-800-1000 s/mm<sup>2</sup> (iii) T1w FFE with enhancement and (iv) sample Dynamic Contrast Enhanced (DCE) curves of the lesion core, defined as red colored ROI in the T1w FFE images. Scale bar 3cm. For MRI sequence parameter details see Table 2.

<https://doi.org/10.1371/journal.pone.0185995.g002>



1. Affine image co-registration
2. Pixel detection step for feature extraction and non linear processing, such as ADC maps from DWI sequences (Fig 2Aii, 2Bii and 2Cii), kinetic parameters of the DCE-profile such as  $K_{trans}$ ,  $V_e$  and  $K_{ep}$  according to Tofts' pharmacokinetic model [24,25] and normalization of T2w images based on a rectangular prostatic reference volume for calculation of first and second order texture features.
3. Feature classifier with 3 logical steps: Step 1 defines the dimension of predictor space and the transformation parameters. Step 2 is a linear summation parameter that integrates step 1 information to construct a scalar map. Step 3 features an error-feedback method, which has been trained by a supervised learning process to achieve a congruence of the scalar value (step 2) with the Gleason grade.

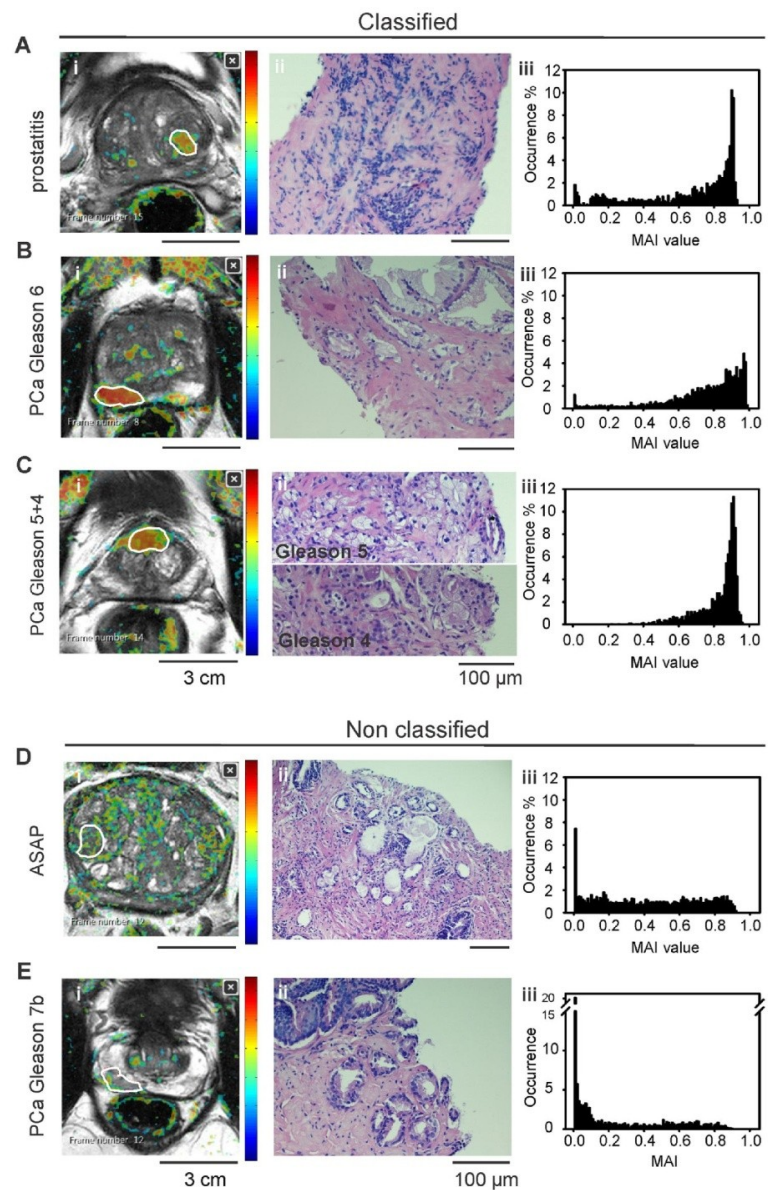
The final output of the feature classifier is a pixel-based malignancy prediction score, ranging from 0 to 1, termed as Malignancy Attention Index (MAI). Watson Elementary™ constructs a malignancy prediction heatmap with high MAI values represented in warm colors. This map is projected onto T2w images, thus anatomically highlighting suspect lesions. Moreover, MAI pixel values of each manually defined lesion are automatically sorted in a histogram (malignancy prediction histogram). Histogram shape, mean and median MAI have been suggested as PCa biomarkers [23].

### Lesion definition and evaluation of the volume of interest

Target lesions, i.e. lesions that have been selected for MRI-guided biopsy based on their PI-RADS score, were evaluated retrospectively, after PI-RADS scoring and after the histological report of the biopsy. Lesions were manually defined in T2w images (for the transitional zone) and ADC maps (for the peripheral prostate zone) according to the biopsy needle position by a third radiologist (IEP) blinded to the PI-RADS scoring and the histological identity to avoid bias in sampling. Data analysis was performed by a medical student (AT) and a radiologist (IEP), both blinded to the PI-RADS scoring and histological identity to avoid bias in lesion classification. Watson Elementary™ allows for definition of consecutive ROIs in 2D planes to form a 3D lesion volume of interest. The system generates a probabilistic heatmap (with the MAI values) for each section, which is projected onto the corresponding T2w image. The predefined ROIs were copied from the imaging series to the probabilistic map and the system generated for each predefined lesion (sum of ROIs) a feature summary in a.pdf including (i) a histogram of MAI values, (ii) the average DCE curve (iii) the lesion volume [26]. User access to the ground data is possible only for the probabilistic histograms for each region and not for the intermediate meta-data such as DCE curves, which are only graphically displayed in.pdf format. The classification of lesions was performed visually, by evaluating (i) whether the heatmap of the lesion is distinguishable from the background and (ii) the skewness of the MAI histogram (Fig 3). Classification was performed by a radiologist (IEP) and a medical student (AT), both blinded to the PI-RADS scoring and histological identity of the lesions. In the first classification step observers made no assumption about the histological identity. After classification, the histological identity was unmasked to calculate the specificity and the sensitivity of the software for malignant lesions.

### Statistics

The Statistical Package for the Social Sciences (SPSS Version 21, IBM, Armonk, NY, USA) was used for statistical analysis and graphical plotting. Data were screened for normality using the Shapiro-Wilk test. Values are expressed as median/IQR (interquartile range) and rounded up



**Fig 3. Automated mpMRI classification and histological ground truth.** Target lesions were manually drawn by human observers. Sample images for (A-C) classified and (D, E) non classified PCa and benign biopsy cores. (A) Prostatitis, (B) adenocarcinoma (PCa) Gleason grade 6, (C) PCa Gleason grade 9, (D) atypical small acinar proliferation (ASAP) and (E) PCa Gleason grade 7. From left to right: (i) T2w with MAI-heatmap and outlined lesions (white line), (ii) Hematoxylin-Eosin histopathology of the corresponding biopsy cores and (iii) MAI histograms. Warm colors in MAI heatmaps (i) represent higher values in a scale 0–1. Classified lesions revealed a “warm-colored” MAI-map and a left-skewed histogram.

<https://doi.org/10.1371/journal.pone.0185995.g003>

to the second decimal place unless otherwise stated. Statistical significance was tested using the t-test or the ranksum Mann-Whitney test for unpaired data. MAI score between groups were compared with Kruskal-Wallis ANOVA on ranks with Dunn's posthoc test. Linear correlations were tested with the Pearson product moment correlation coefficient. Continuous probability distribution, as well as independence of nominal data, was tested by means of the chi-squared test. ROC curves were calculated for MAI median, MAI mean and MAI median-to-mean ratio as skewness index as well as for ADC values and PI-RADS reading scores. Statistical significance was set at  $P < 0.05$ .

## Results

### Automated feature extraction: Detection of the arterial input function

The first step in image processing by Watson Elementary™ is feature extraction. For the DCE kinetic analysis, the tissue signal is normalized on the perfusion curve of the common femoral artery (arterial input function). Arterial detection is semi-automated and has to be manually confirmed. The correct artery position was automatically defined in 61 (77.22%) patients; in the remaining 18 (22.78%), it had to be manually reassessed.

### CAD-sensitivity and specificity

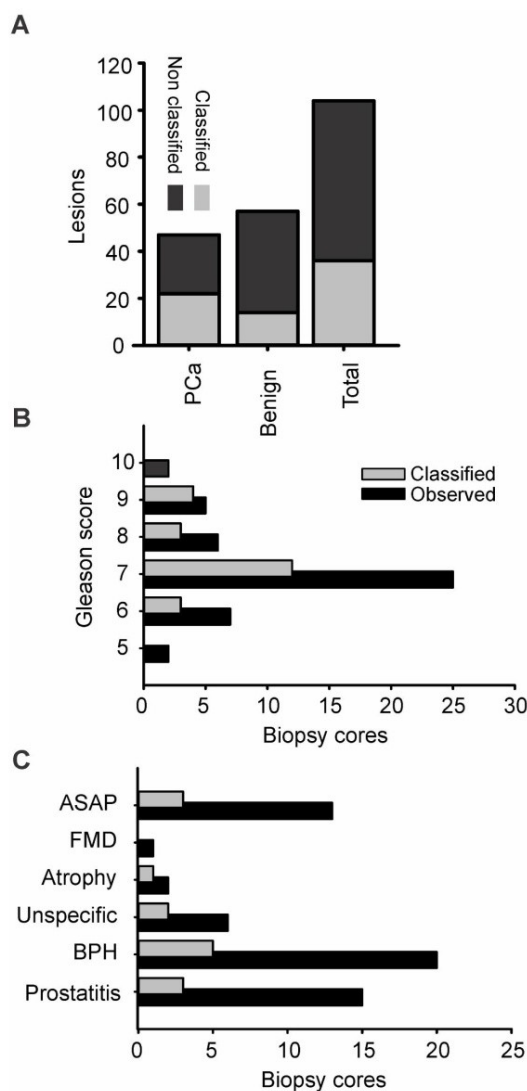
All 104 biopsy cores from 79 patients were scored according to PI-RADS™ in the initial diagnostic dataset (Table 1). Datasets acquired prior to the establishment of PI-RADS™ v2 were re-scored post hoc so that PI-RADS™ v1 and -v2 scores were available for all included cores. Switching between PI-RADS versions, as a consequence, over- or underscored some lesions. Although PI-RADS 2 lesions were not subjected to a biopsy, some were included in the analyzed database. Those lesions were biopsied as PI-RADS 3 or 4 according to PI-RADS™ v1 and then underscored to PI-RADS 2 after the introduction of PI-RADS™ v2 (Table 1).

In the joined MAI/T2w images (Fig 3A–3E, left panels), we evaluated whether the heatmap of our target lesions was visually distinguishable from the background, i.e. whether the lesion could be detected in the MAI/T2w heatmap without the use of ADC maps. Warm-colored ROIs with a characteristic left-skewed MAI histogram were evaluated as classified (detected) lesions (Fig 3A–3C), in sharp contrast to the flat or right-skewed MAI histogram shape of the non classified (cold-colored) ROIs (Fig 3D and 3E). In previous work, a MAI-max cut-off value of 0.6 was used as criterion for malignant lesion classification, based on the assumption that MAI linearly correlates with the Gleason score [23]; this method was not applicable in our study because almost all lesions, regardless of their histogram shape and histological identity, showed a maximum MAI value higher than 0.6 (Supporting information, S1).

From 47 histologically confirmed malignant (PCa) and 57 benign lesions, 22 PCa and 14 benign lesions were classified (Fig 4A and Table 3). The deduced CAD-sensitivity for prostate malignancy in our series was 46.81% with a specificity of 75.44% and a PPV of 61.11% (Table 4).

The hypothesis whether CAD preferentially classifies malignant over benign cores was tested with a chi-squared test, which revealed independent classification of benign and malignant biopsy cores, albeit with a slight trend towards the classification of malignancy,  $P 0.06$  (Table 3). Furthermore, we questioned whether CAD sensitivity leaned towards a particular histological identity, a particular Gleason malignancy grade or a specific benign condition. All malignant lesions ( $n 47$ ) were typified as acinar adenocarcinoma of various Gleason grades from 5 to 10. As graphically demonstrated in (Fig 4B), CAD-sensitivity was not influenced by the malignancy grade,  $P 0.713$   $R 0.193$  Pearson correlation. Among false positives (i.e. benign cores falsely classified) (Fig 4C), 5/14 lesions (35.71%) corresponded to benign prostate





**Fig 4. Classification of malignant and benign biopsy cores.** (A) Proportion of classified to non classified malignant and benign lesions. From 104 mixed type cores 36 were classified, from those (classified/total) 14/57 were benign and 22/47 malignant. (B) The classification was independent of the Gleason grade,  $P = 0.713$  R 0.193 Pearson correlation (C) Classification of benign lesions. ASAP, Atypical Small Acinar Proliferation; FMD, Fibromuscular Dysplasia; BPH, Benign Prostatic Hyperplasia.

<https://doi.org/10.1371/journal.pone.0185995.g004>

**Table 3. Identity of classified and non classified lesions.**

Patient <i>N</i> = 79	Classified	Non-classified	Total
Malignant (PCa)	22	25	47
Benign	14	43	57
Total	36	68	104

*P* = 0.06, Chi square test

The classified malignant cores are the true positives (TP), the non classified malignant cores are sorted as false negatives (FN), the classified benign cores as false positives (FP) and the non classified benign cores as true negatives (TN). Chi square test returns equal occurrence of the outcomes for the two groups, *P* 0.06,  $\chi^2$  test.

<https://doi.org/10.1371/journal.pone.0185995.t003>

hyperplasia, followed by prostatitis and atypical small acinar proliferation with 3/14 lesions, 21.43% each.

### Malignancy attention index as biomarker

In a previous study [23], MAI was proposed as a potential PCa biomarker and MAI heatmaps/histograms as potential core malignancy profiles (Fig 3).

In the database used in our study, malignant lesions showed the expected ADC dip compared to benign ones, with an ADC (median/IQR) of 650/295 ( $\times 10^{-6}$ ) mm<sup>2</sup>/s for malignant and 950/340 ( $\times 10^{-6}$ ) mm<sup>2</sup>/s for benign foci, *P* < 0.001 Mann-Whitney U-test (Fig 5A). This is important because ADC values are widely accepted as the most relevant malignancy feature, especially for the peripheral zone, and are thus highly ranked by CAD-classifiers [27–32]. Median MAI values were selected as the most representative descriptive parameter of skewed histograms. The comparison between all observed (classified and non classified) benign and malignant biopsy cores indeed revealed a significantly lower MAI score for benign lesions, with (median/IQR) 0.39/0.18 compared to malignant ones 0.56/0.24, *P* 0.023 Mann-Whitney U-test (Fig 5B).

MAI score was, as expected, significantly higher in classified compared to non classified lesions regardless of identity, *P* < 0.05 Kruskal-Wallis ANOVA on ranks with Dunn's post hoc test. However, classified PCa and benign cores did not show any significant MAI difference, *P* < 0.05 Kruskal Wallis ANOVA on ranks with Dunn's post hoc test (Fig 5C).

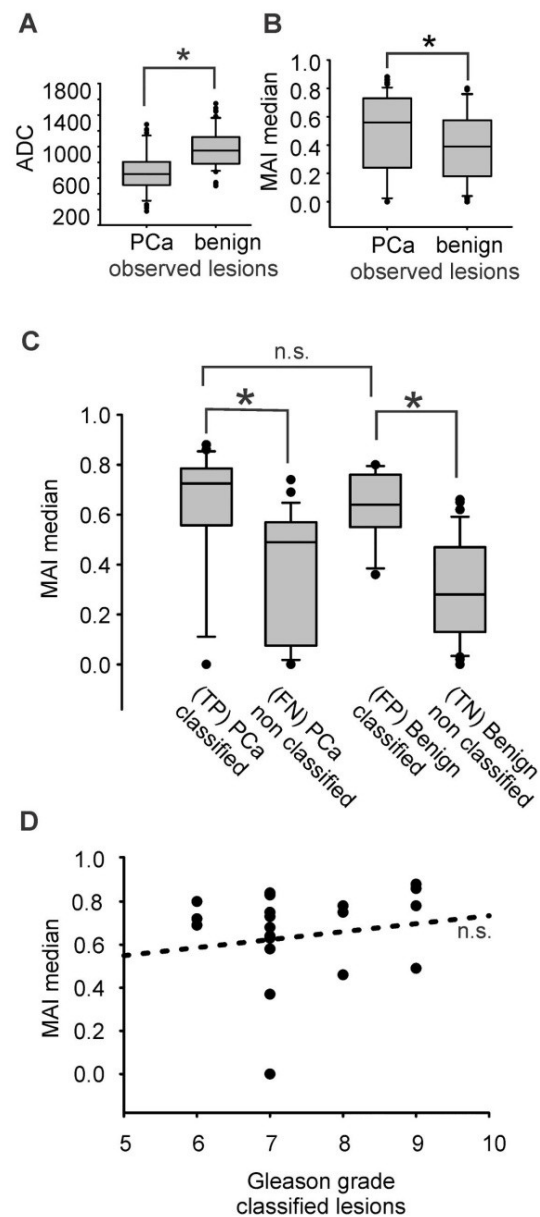
Furthermore, we tested whether MAI score qualifies for a Gleason's grade predictor using the Pearson's test. Within the 22 classified PCa biopsy cores, MAI did not show any significant correlation with Gleason grade, *P* 0.52 *R* 0.14 Pearson product moment correlation (Fig 5D).

ROC analysis of median and mean MAI as malignancy predictors revealed rather poor results with an area under curve  $\pm$  standard error of the mean (AUC $\pm$ SEM) 0.63 $\pm$ 0.06 (95% CI 0.52–0.74), *P* 0.02 for MAI median and 0.64 $\pm$ 0.06 (95% CI 0.53–0.75), *P* 0.02 for MAI mean. The predictive outcome of the median/mean ratio as skewness index was not significant with

**Table 4. CAD sensitivity and specificity for PCa detection.**

False negative rate (FNR)	0.53 lesions/patient
Sensitivity	46.81%
Specificity	75.44%
Positive predictive value (PPV)	61.11%
Negative predictive value (NPV)	63.24%
False positive rate (FPR)	24.56%
False detection rate (FDR)	38.89%

<https://doi.org/10.1371/journal.pone.0185995.t004>



**Fig 5. MAI as malignancy biomarker and MAI correlation with the Gleason grade.** (A) Average ADC values ( $\times 10^{-6} \text{ mm}^2/\text{s}$ ) of observed lesions, 650/295 (median/IQR) for malignant and 950/339.5 for benign,  $N$  (malignant/benign) 47/57,  $P < 0.0001$  Mann-Whitney U-test (B) MAI of observed lesions, 0.56/0.23 (median/IQR) for malignant and 0.39/0.18 for benign lesions,  $N$  (malignant/benign) 47/57,  $P 0.023$  Mann-Whitney U-test. (C) MAI in benign, malignant, classified and non classified biopsy cores,  $P < 0.001$  Kruskal-Wallis

ANOVA on ranks, \*  $P < 0.05$  Dunn's post hoc test (D) No significant correlation between MAI and Gleason grade,  $P = 0.522$ ,  $N = 22$ ,  $R = 0.144$ , Pearson's product moment correlation; n.s., non significant.

<https://doi.org/10.1371/journal.pone.0185995.g005>

AUC  $0.59 \pm 0.06$  (95% CI 0.47–0.70),  $P = 0.12$  (Fig 6, Table 5). MAI did not significantly improve the board certified reader's accuracy using the American College of Radiology (ACR) and European Society for Uradiology (ESUR) standards PI-RADS™ v1 (AUC  $0.67 \pm 0.05$  with 95% CI 0.58–0.76,  $P = 0.003$ ) and PI-RADS™ v2 (AUC  $0.68 \pm 0.04$  with 95% CI 0.59–0.76,  $P = 0.002$ ). The optimal MAI median cut-off point estimated with Youden statistics was 0.54 with a sensitivity of 61.7% (95% CI 46.38–75.49%) and specificity of 68.42% (95% CI 54.76–80.09%) (Fig 6 and Table 5). By setting an optimized cut-off point for MAI mean, however, we could improve the sensitivity and specificity up to 70.21% / 61.4% (95% CI 55.11–82.66% and 47.57–74%, respectively) (Fig 6 and Table 5). Moreover, analysis of the ADC value alone showed a stronger predictive behavior compared to the software-calculated MAI with AUC  $0.79 \pm 0.05$  (95% CI 0.70–0.88) and  $P = 0.04$  compared to MAI, chi-squared test (Fig 6).

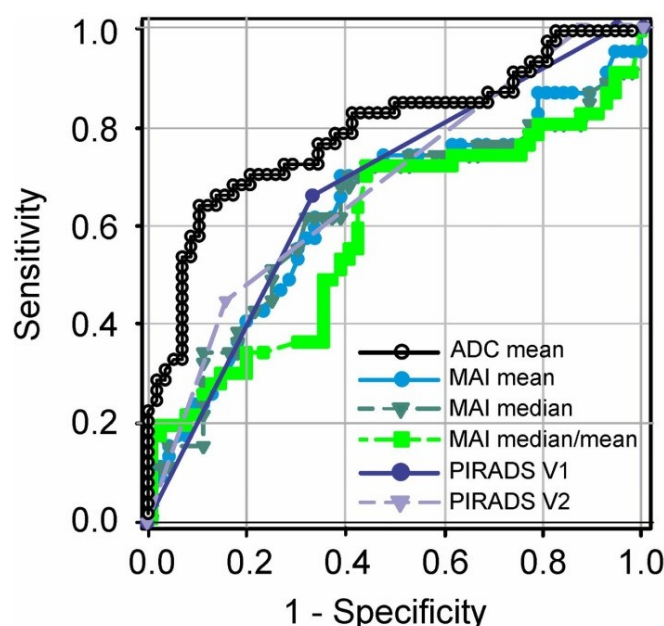
Guided by the hypothesis that Watson Elementary™ might be more specific for particular lesion locations and sizes, we tested for possible predilection towards the peripheral or the central zone of the prostate gland. In the transitional prostate zone CAD has classified 10 out of 25 histologically confirmed PCa; whereas in the peripheral zone, 12 out of 22, hence with no apparent influence on the performance ( $P = 0.481$  chi-squared test). The lesion volume, however, had a significant influence on the CAD-performance. Amongst lesions smaller than 0.5ml (Fig 7A) the vast majority was not classified (sensitivity 27.27% and FNR 31.37%). For intermediate size cores of 0.5ml–1.0ml, the CAD revealed an improved performance (sensitivity 53.33% and FNR 18.42%) and false negatives were minimized for lesions larger than 1.0 ml (sensitivity 80%, FNR 13.33%). As expected, lesion volume was independent of Gleason grade,  $R = 0.18$   $P = 0.15$  Pearson's correlation. It's worth noticing that it is more crucial to eliminate the number of FN than the number of FP because the therapeutic consequence for the patient would be an undiagnosed PCa in the first case, compared to an unnecessary biopsy in the second case. In this context, CAD-performance is satisfactory for lesions larger than 1.0ml (Fig 7A). In Fig 7B, the MAI score of classified and non classified cores is plotted with the lesion volume. There is a strong trend for a positive correlation between lesion size and MAI-score for TP lesions ( $P = 0.057$  Pearson's correlation) but not for any other category (TN, FP and FN,  $P > 0.1$  Pearson's correlation). We questioned the clinical significance of lesions smaller than 0.5 ml, which make up 49.04% of our database. Interestingly, the malignancy incidence between lesions smaller than 0.5 ml, and those that were larger, was identical (Fig 7Ci and 7Ciii) with approximately 43% probability of malignancy in both groups. Moreover, within malignant lesions, we observed comparable PCa aggressiveness in terms of Gleason grade (Fig 7Cii and 7Civ), with high-grade cancers being equally possible in both small and larger lesions.

In summary, MAI is a weak PCa biomarker, especially for lesions smaller than 0.5ml in either the transitional and peripheral zone, regardless of lesion aggressiveness.

## Discussion

This study aims to emphasize the growing necessity for commercialized prostate mpMRI CAD-software tools for the radiological, and perhaps urological, praxis. By retrospectively testing 104 lesions (47 malignant, 57 benign) in a series of 79 patients, a commercialized prostate CAD, Watson Elementary™, revealed a sensitivity of 46.81% for prostate malignancy, with a specificity of 75.44% and a PPV of 61.11%. Our results considerably differ from previous reports on the same software. Roethke et al. [23] have tested Watson Elementary™ in a cohort of 45 patients with 1102 MR/TRUS acquired biopsy cores (76 malignant/1026 benign) and





**Fig 6. Receiver operating characteristic (ROC) trade-off curve for ADC, MAI and PI-RADS.** ADC mean (black), MAI median (dark green), MAI mean (light blue) and MAI median/mean ratio (green). MAI is compared with the ADC performance alone and with the observer's performance according to PI-RADS™ v1 (dark blue) and PI-RADS™ v2 (violet). The area under the curve (AUC) is  $0.64 \pm 0.057$  (mean, SEM) with 95% CI 0.53–0.75 and  $P = 0.02$  for MAI mean,  $0.63 \pm 0.058$  with 95% CI 0.52–0.74 and  $P = 0.02$  for MAI median and  $0.59 \pm 0.058$  with 95% CI 0.47–0.70 and  $P = 0.13$  for the MAI median/mean ratio. Corresponding values for the mean ADC lesion value are AUC  $0.79 \pm 0.05$  with 95% CI 0.70–0.88,  $P < 0.0001$ . Observer's performance for PI-RADS™ v1 was AUC  $0.67 \pm 0.05$  with CI 0.58–0.76,  $P = 0.003$  and for PI-RADS™ v2 AUC  $0.68 \pm 0.04$  with CI 0.59–0.76,  $P = 0.002$ .  $N$  malignant/benign cores 47/57. MAI and PI-RADS (v1, v2) reveal comparable performances in malignancy detection,  $P = 0.60$  for MAI vs PI-RADS v1 and  $P = 0.53$  for MAI vs PI-RADS v2, chi-squared test. ADC is superior to MAI in malignancy prediction,  $P = 0.04$ , chi-squared test.

<https://doi.org/10.1371/journal.pone.0185995.g006>

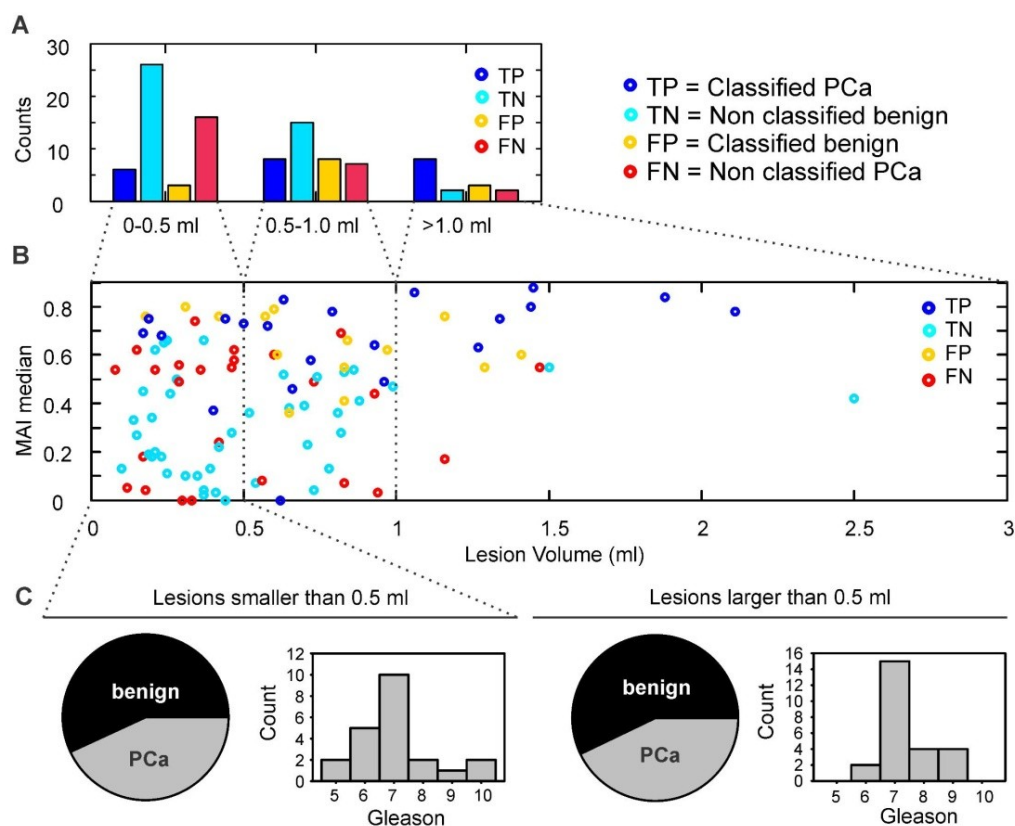
**Table 5. Receiver operating characteristic (ROC) analysis for MAI and PI-RADS.**

	Area under the curve (AUC)				<i>P</i>	Cut-off	Sensitivity at cut-off			Specificity at cut-off		
	AUC	SEM	95% CI				Sensitivity	95% CI		Specificity	95% CI	
MAI mean	0.64	0.06	0.53	0.75	0.02	0.435	70.21	55.11	82.66	61.4	47.57	74
MAI median	0.63	0.06	0.47	0.7	0.02	0.535	61.7	46.38	75.49	68.42	54.76	80.09
PIRADS v1	0.67	0.05	0.58	0.76	0.003							
PIRADS v2	0.68	0.04	0.59	0.76	0.002							

Sensitivity and specificity for MAI and PI-RADS as estimated with Receiver Operating Characteristic (ROC) curve analysis and Youden statistics. MAI mean and MAI median correspond to the mean and median values of the MAI histogram describing each lesion. PI-RADS v1 and -v2 are evaluated as malignancy prediction indices compared to MAI.  $N$  malignant/benign cores 47/57. AUC, Area Under Curve; SEM, Standard Error of Mean; CI, Confidence Interval; PI-RADS, Prostate Imaging Reporting And Diagnosis System.

<https://doi.org/10.1371/journal.pone.0185995.t005>





**Fig 7. CAD performance is reliable for lesions larger than 1 ml.** (A) Counts of classified and non-classified cores in relation to their volume. Volume distribution of all lesions (number, %): 0–0.5ml (51, 49.04), 0.5–1.0ml (38, 36.54) and larger than 1.0ml (15, 14.42). True positive (TP, blue) lesions include classified PCa, true negatives (TN, cyan) are the non-classified benign cores, false positives (FP, yellow) are the classified benign cores and false negatives (FN, red) the non-classified PCa. The CAD-sensitivity increases and the number of FN decreases towards larger lesion volumes: (sensitivity % / FNR %) 27.27/31.37 for 0–0.5ml lesions, 53.33/18.42 for 0.5–1.0ml lesions and 80.00/13.33% for lesions larger than 1.0ml. (B) MAI score with lesion volume. A strong trend for a positive correlation between lesion size and MAI-score was found for TP lesions ( $P = 0.057$ , Pearson's correlation) but not for any of the remaining categories (TN, FP and FN,  $P > 0.1$ , Pearson's correlation). (C) Lesions smaller than 0.5 ml show the same malignancy incidence and comparable aggressiveness compared to larger lesions (Ci) Lesions smaller than 0.5 ml (number, %) benign (29, 56.86) malignant (22, 43.14), (Cii) Gleason histogram for malignant lesions smaller than 0.5 ml, (Ciii) Lesions larger than 0.5 ml (number, %) benign (28, 57.14) malignant (21, 42.86), (Civ) Gleason histogram for malignant lesions larger than 0.5 ml.

<https://doi.org/10.1371/journal.pone.0185995.g007>

achieved a sensitivity of 85.71% and specificity of 87.50% when setting an optimal MAI mean cut-off threshold for malignancy detection. By setting an optimized cut-off value of MAI mean in our study, we could improve the sensitivity and specificity up to 70.21%/61.4%, which is inferior to the previously reported values but comparable in terms of methodology [23]. This considerably differs from previous promising studies that have established custom-made software tools for mpMRI analysis with high stand-alone accuracy for malignancy detection [33–36]. The group of Litjens et al. [26] achieved a stand-alone accuracy of AUC = 0.89 in a remarkably large database of 347 patients.

The outcome discrepancy between our series and previous testing of the same software [23] could be attested to a variety of causes. In terms of methodology, a previous study applied a combination of systematic and MRI-guided transperineal biopsies, ending up with more probes per patient (approximately 25) compared to the current study, which was based exclusively on MRI-guided transrectal biopsies (approximately 2 probes per lesion) [23]. Nevertheless, regardless of the number of biopsies per patient/core, both studies define a confirmed lesion by at least one positive biopsy. Roethke et al. [23] report their results in patient-based percentages, in contrast to our study which is needle-based. Taking into account the different reporting methods, our results are technically comparable with the needle-based results of Roethke et al., i.e. sensitivity/specificity 54.67%/97.76%.

Another methodological variation that might have influenced the discrepancy from previous work is that Roethke et al. implement a MAI-max cut-off value of 0.6 as criterion for malignant lesion classification [23]. This method, though more objective than visual classification, was not applicable in our study because almost all lesions showed a maximum MAI value higher than 0.6 (Supporting information, S1).

A prerequisite of the high CAD-accuracy is the training of the classifier on a database with similar characteristics to the testing database [13]. An important limitation of this study is the lack of interaction with the classifier of the commercially available tested CAD-software [23]. Despite the classifier having been trained on scanner data with the same (3T) field strength, factors such as the different technical characteristics, coils, static magnetic field inhomogeneities and protocols for the resonance frequency adjustment led to contrast differences that could sufficiently affect the outcome. In a thorough review by Wang et al. [13], numerous studies with databases varying between 15 and 100 patients were compared, not only in terms of performance but also in terms of the analyzed modalities, field strength, ground truth, the method for candidate lesion generation and applied classifier, revealing a broad heterogeneity. Implementation of different receiver coils, such as the use of endorectal coil [33,34], increases the methodological variation.

Moreover, variations in the applied DWI b-values, i.e. b 0–800 s/mm<sup>2</sup> in a previous study [23] compared to b 0–1000 s/mm<sup>2</sup> in our study, could affect the ADC computation and inject a significant classification error probability [32,34]. ADC values are widely accepted as the most relevant malignancy feature and rank highly in CAD-classifiers [27–32]. Interestingly, previous reports [37–40] have suggested the superiority of b values of 1000 and 1500 s/mm<sup>2</sup> compared to either 300, 500 or 2000 s/mm<sup>2</sup>.

The lesion volume should be considered as an independent factor. The current study had, as a single exclusion criterion, technical data incompatibility with Watson Elementary™. In ca. 86% of the cases, the sampled lesion's volume was smaller than 1.0 ml; and in 49%, smaller than 0.5 ml, in keeping with early PCa diagnosis. Watson Elementary™ showed promising performance only in lesions larger than 1.0 ml, which might explain differences with previous studies where volume inclusion criteria might have differed. In the studies of Roethke et al. [23,41], lesions smaller than 0.5 ml were not considered clinically significant cancer, in line with ESUR guidelines 2012 [17]. However, lack of methodological definition on volume selection criteria does not allow for a more elaborated comparison. It is remarkable that both the malignancy and Gleason grade of smaller, „clinically insignificant“ lesions do not differ compared to larger lesions, as shown in Fig 7C. This result supports the existing body of evidence that small cancers can significantly affect a patient's outcome and encourages biopsy and treatment according to guidelines [42].

Furthermore, lack of access to whole-mount prostate pathology was a limitation of this study with possible influence on the results. The classifier of Watson Elementary™ has been regularized to create congruence with malignancy grade, a.k.a. Gleason grade [23]. Nonetheless,

MAI did not significantly correlate with the pathological outcome, which is the Gleason score in our database. Inter-observer differences in Gleason grading between the training and the testing database could already contribute to this discrepancy, albeit minimally, as previous studies have shown negligible inter-observer variation mostly at the upper and lower limit, namely 4 and 8–10 of the Gleason scale [43,44]. Another possible variability factor may rely on the classifier's training on whole slide pathology specimens, whereas the grading process in the current study was based on needle biopsies [45,46]. Moreover, the unequal sample distribution with the majority of patients revealing Gleason 6 or 7 (3+3 or 3+4) at the time of diagnosis may bias Pearson's correlation coefficient negatively [47]. However, such an unequally weighted Gleason distribution is the typical occurrence pattern in population screening and should be taken into account in the training process of a detection method [48,49].

The ROC—analysis of the observer's performance reveals a reduced accuracy of our study compared with previously published results on PI-RADS [16,41]. Kasel-Seibert et al. In a recent study, compared the performance of PI-RADS version 2 to version 1, showing a high accuracy of AUC of 0.88 and 0.91 for v1 and v2, respectively, in the hands of experienced MRI readers [50]. However, considerable differences in the methodology and database selection should be taken into consideration. Patients with PI-RADS 1 or 2 were not subjected to biopsy and therefore not included in our study. The majority of PI-RADS 3 patients was also not biopsied in our hospital, and were therefore excluded from this study. A limited number of included PI-RADS 2 and 3 lesions derived mostly by downgrading PI-RADS v1 4 lesions in the re-evaluation process after the introduction of PI-RADS v2. On the other hand, Kasel-Seibert et al. [50], as well as other previous studies [26] included PI-RADS 3 lesions in their official selection criteria. Another considerable bias-introducing factor, as the authors also acknowledge, is the patient selection criteria (systematic biopsy was not performed) and the relative low (29%) malignancy rate within the selected population [50]. In the current study, almost all patients were subjected to a (non conclusive or negative) systematic biopsy and the malignancy rate was ca. 45%, thus considerably higher compared to previously published data [50].

Recently, studies that evaluate the role of CAD implementation in improving the radiologist's performance have been gaining ground on those evaluating stand-alone performance, such as our study [51,52]. Large scale approaches ( $n = 89$  [51] and  $n = 107$  [52]) implemented different methodologies to show that CAD implementation improved radiologists' sensitivity from 80.9% to 87.6% [51]. Accuracy without and with CAD-combined reading for differentiation between benign and malignant (AUC 0.81 versus 0.88), indolent and aggressive lesions (AUC 0.78 versus 0.88) was improved, respectively [52]. In both original research works, the CAD classifier has been previously established and trained on a comparable database, in contrast to our study where interaction with the classifier was not possible.

This study shows that a carefully designed commercialized CAD software (Watson Elementary™) does not perform satisfactorily when tested with a different instrumentation and imaging configuration, despite using almost double the number of patients compared to previous studies [13,23,33–35]. Lack of whole-mount prostate pathology, the low number of PI-RADS 2 and 3 lesions and the challenging character of the database including small lesions, not necessarily encountered as significant in previous studies, should be encountered as possible limiting factors. It is worth mentioning that the scanning parameters applied in our department fulfill the recommendations for diagnosis as defined by the American College of Radiology in PI-RADS™ v2. In line with previous observations reviewed by Wand et al. [13], the results of this study support that super-optimistic CAD-performances might be dataset-bound. Altogether, while being in the right framework, the tested software is not satisfactory yet. A necessary requirement of a CAD-software is the ability to apply and generalize to different scanning



settings. A broader, optimally multicenter pool of datasets for broader and maybe interactive classifier training should be implemented to improve the general applicability of CAD systems [13].

## Supporting information

**S1 File. Original data and metadata in.xlsx spreadsheets.**  
(XLSX)

## Acknowledgments

Authors thank Joachim Feger, Claudia Kurat and Arnhold Kott for PI-RADS scoring as board-certified radiologists. Rachel Wong deserves special acknowledgements for proofreading the manuscript as native English speaker.

## Author Contributions

**Conceptualization:** Ansgar Malich, Ismini Papageorgiou.

**Data curation:** Cornelia Tennstedt-Schenk, Sven Winzler, Ansgar Malich, Ismini Papageorgiou.

**Formal analysis:** Anika Thon.

**Investigation:** Anika Thon, Sven Winzler, Ansgar Malich, Ismini Papageorgiou.

**Methodology:** Ansgar Malich, Ismini Papageorgiou.

**Project administration:** Ansgar Malich, Ismini Papageorgiou.

**Resources:** Cornelia Tennstedt-Schenk.

**Supervision:** Ulf Teichgräber, Ansgar Malich.

**Validation:** Stathis Hadjidemetriou.

**Visualization:** Anika Thon, Ismini Papageorgiou.

**Writing – original draft:** Anika Thon, Stathis Hadjidemetriou, Ansgar Malich, Ismini Papageorgiou.

**Writing – review & editing:** Ulf Teichgräber, Ansgar Malich.

## References

1. Cancer of the Prostate—SEER Stat Fact Sheets [Internet]. [cited 19 Oct 2016]. <http://seer.cancer.gov/statfacts/html/prost.html>
2. Prostate cancer survival statistics. In: Cancer Research UK [Internet]. 15 May 2015 [cited 19 Oct 2016]. <http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer/survival>
3. Filella X, Foj L. Prostate Cancer Detection and Prognosis: From Prostate Specific Antigen (PSA) to Exosomal Biomarkers. *Int J Mol Sci*. 2016; 17. <https://doi.org/10.3390/ijms17111784> PMID: 27792187
4. Ilic D, Neuberger MM, Djulbegovic M, Dahm P. Screening for prostate cancer. *Cochrane Database Syst Rev*. 2013; CD004720. <https://doi.org/10.1002/14651858.CD004720.pub3> PMID: 23440794
5. Pron G. Prostate-Specific Antigen (PSA)-Based Population Screening for Prostate Cancer: An Evidence-Based Analysis. *Ont Health Technol Assess Ser*. 2015; 15: 1–64.
6. Hayes JH, Barry MJ. Screening for prostate cancer with the prostate-specific antigen test: a review of current evidence. *JAMA*. 2014; 311: 1143–1149. <https://doi.org/10.1001/jama.2014.2085> PMID: 24643604

7. Wolf AMD, Wender RC, Etzioni RB, Thompson IM, D'Amico AV, Volk RJ, et al. American Cancer Society Guideline for the Early Detection of Prostate Cancer: Update 2010. *CA Cancer J Clin*. 2010; 60: 70–98. <https://doi.org/10.3322/caac.20066> PMID: 20200110
8. Mottet N, Bellmunt J, Bolla M, Briers E, Cumberbatch MG, De Santis M, et al. EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. *Eur Urol*. 2016; <https://doi.org/10.1016/j.eururo.2016.08.003> PMID: 27568654
9. Bjurlin MA, Meng X, Le Nobin J, Wysock JS, Lepor H, Rosenkrantz AB, et al. Optimization of prostate biopsy: the role of magnetic resonance imaging targeted biopsy in detection, localization and risk assessment. *J Urol*. 2014; 192: 648–658. <https://doi.org/10.1016/j.juro.2014.03.117> PMID: 24769030
10. Oberlin DT, Casalino DD, Miller FH, Meeks JJ. Dramatic increase in the utilization of multiparametric magnetic resonance imaging for detection and management of prostate cancer. *Abdom Radiol N Y*. 2016; <https://doi.org/10.1007/s00261-016-0975-5> PMID: 27858090
11. Klotz L, Emberton M. Management of low risk prostate cancer-active surveillance and focal therapy. *Nat Rev Clin Oncol*. 2014; 11: 324–334. <https://doi.org/10.1038/nrclinonc.2014.73> PMID: 24821214
12. Chang AJ, Autio KA, Roach M, Scher HI. High-risk prostate cancer-classification and therapy. *Nat Rev Clin Oncol*. 2014; 11: 308–323. <https://doi.org/10.1038/nrclinonc.2014.68> PMID: 24840073
13. Wang S, Burt K, Turkbey B, Choyke P, Summers RM. Computer Aided-Diagnosis of Prostate Cancer on Multiparametric MRI: A Technical Review of Current Research. *BioMed Res Int*. 2014; 2014: e789561. <https://doi.org/10.1155/2014/789561> PMID: 25525604
14. Schiebler ML, Schnall MD, Pollack HM, Lenkinski RE, Tomaszewski JE, Wein AJ, et al. Current role of MR imaging in the staging of adenocarcinoma of the prostate. *Radiology*. 1993; 189: 339–352. <https://doi.org/10.1148/radiology.189.2.8210358> PMID: 8210358
15. Hamoen EHJ, de Rooij M, Witjes JA, Barentsz JO, Rovers MM. Use of the Prostate Imaging Reporting and Data System (PI-RADS) for Prostate Cancer Detection with Multiparametric Magnetic Resonance Imaging: A Diagnostic Meta-analysis. *Eur Urol*. 2015; 67: 1112–1121. <https://doi.org/10.1016/j.eururo.2014.10.033> PMID: 25466942
16. Kayat Bittencourt L, Litjens G, Hulsbergen-van de Kaa CA, Turkbey B, Gasparetto EL, Barentsz JO. Prostate Cancer: The European Society of Urogenital Radiology Prostate Imaging Reporting and Data System Criteria for Predicting Extraprostatic Extension by Using 3-T Multiparametric MR Imaging. *Radiology*. 2015; 276: 479–489. <https://doi.org/10.1148/radiol.15141412> PMID: 25848900
17. Barentsz JO, Richenberg J, Clements R, Choyke P, Verma S, Villeirs G, et al. ESUR prostate MR guidelines 2012. *Eur Radiol*. 2012; 22: 746–757. <https://doi.org/10.1007/s00330-011-2377-y> PMID: 22322308
18. Haider MA, Chung P, Sweet J, Toi A, Jhaveri K, Ménard C, et al. Dynamic contrast-enhanced magnetic resonance imaging for localization of recurrent prostate cancer after external beam radiotherapy. *Int J Radiat Oncol Biol Phys*. 2008; 70: 425–430. <https://doi.org/10.1016/j.ijrobp.2007.06.029> PMID: 17881141
19. Kozlowski P, Chang SD, Jones EC, Berean KW, Chen H, Goldenberg SL. Combined diffusion-weighted and dynamic contrast-enhanced MRI for prostate cancer diagnosis—correlation with biopsy and histopathology. *J Magn Reson Imaging JMRI*. 2006; 24: 108–113. <https://doi.org/10.1002/jmri.20626> PMID: 16767709
20. Kirkham APS, Emberton M, Allen C. How good is MRI at detecting and characterising cancer within the prostate? *Eur Urol*. 2006; 50: 1163–1174; discussion 1175. <https://doi.org/10.1016/j.eururo.2006.06.025> PMID: 16842903
21. Speight JL, Roach M. Advances in the treatment of localized prostate cancer: the role of anatomic and functional imaging in men managed with radiotherapy. *J Clin Oncol Off J Am Soc Clin Oncol*. 2007; 25: 987–995. <https://doi.org/10.1200/JCO.2006.10.3218> PMID: 17350948
22. Tsivian M, Gupta RT, Tsivian E, Qi P, Mendez MH, Abern MR, et al. Assessing clinically significant prostate cancer: Diagnostic properties of multiparametric magnetic resonance imaging compared to three-dimensional transperineal template mapping histopathology. *Int J Urol Off J Jpn Urol Assoc*. 2016; <https://doi.org/10.1111/iju.13251> PMID: 27859637
23. Roethke MC, Kuru TH, Mueller-Wolf MB, Agterhuis E, Edler C, Hohenfellner M, et al. Evaluation of an Automated Analysis Tool for Prostate Cancer Prediction Using Multiparametric Magnetic Resonance Imaging. *PloS One*. 2016; 11: e0159803. <https://doi.org/10.1371/journal.pone.0159803> PMID: 27454770
24. Tofts PS, Brix G, Buckley DL, Evelhoch JL, Henderson E, Knopp MV, et al. Estimating kinetic parameters from dynamic contrast-enhanced T(1)-weighted MRI of a diffusable tracer: standardized quantities and symbols. *J Magn Reson Imaging JMRI*. 1999; 10: 223–232. PMID: 10508281
25. Parker GJ, Tofts PS. Pharmacokinetic analysis of neoplasms using contrast-enhanced dynamic magnetic resonance imaging. *Top Magn Reson Imaging TMRI*. 1999; 10: 130–142. PMID: 10551628

26. Litjens G, Debats O, Barentsz J, Karssemeijer N, Huisman H. Computer-aided detection of prostate cancer in MRI. *IEEE Trans Med Imaging*. 2014; 33: 1083–1092. <https://doi.org/10.1109/TMI.2014.2303821> PMID: 24770913
27. Hambrock T, Somford DM, Huisman HJ, van Oort IM, Witjes JA, Hulsbergen-van de Kaa CA, et al. Relationship between apparent diffusion coefficients at 3.0-T MR imaging and Gleason grade in peripheral zone prostate cancer. *Radiology*. 2011; 259: 453–461. <https://doi.org/10.1148/radiol.11091409> PMID: 21502392
28. Vos EK, Kobus T, Litjens GJS, Hambrock T, Hulsbergen-van de Kaa CA, Barentsz JO, et al. Multiparametric Magnetic Resonance Imaging for Discriminating Low-Grade From High-Grade Prostate Cancer. *Invest Radiol*. 2015; 50: 490–497. <https://doi.org/10.1097/RLI.0000000000000157> PMID: 25867656
29. Turkbey B, Shah VP, Pang Y, Bernardo M, Xu S, Kruecker J, et al. Is apparent diffusion coefficient associated with clinical risk scores for prostate cancers that are visible on 3-T MR images? *Radiology*. 2011; 258: 488–495. <https://doi.org/10.1148/radiol.10100667> PMID: 21177390
30. Zelhof B, Pickles M, Liney G, Gibbs P, Rodrigues G, Kraus S, et al. Correlation of diffusion-weighted magnetic resonance data with cellularity in prostate cancer. *BJU Int*. 2009; 103: 883–888. <https://doi.org/10.1111/j.1464-410X.2008.08130.x> PMID: 19007373
31. Gibbs P, Liney GP, Pickles MD, Zelhof B, Rodrigues G, Turnbull LW. Correlation of ADC and T2 measurements with cell density in prostate cancer at 3.0 Tesla. *Invest Radiol*. 2009; 44: 572–576. <https://doi.org/10.1097/RLI.0b013e3181b4c10e> PMID: 19692841
32. Hegde JV, Mulkern RV, Panych LP, Fennessy FM, Fedorov A, Maier SE, et al. Multiparametric MRI of Prostate Cancer: An Update on State-of-the-Art Techniques and Their Performance in Detecting and Localizing Prostate Cancer. *J Magn Reson Imaging JMIR*. 2013; 37: 1035–1054. <https://doi.org/10.1002/jmri.23860> PMID: 23606141
33. Shah V, Turkbey B, Mani H, Pang Y, Pohida T, Merino MJ, et al. Decision support system for localizing prostate cancer based on multiparametric magnetic resonance imaging. *Med Phys*. 2012; 39: 4093–4103. <https://doi.org/10.1118/1.4722753> PMID: 22830742
34. Peng Y, Jiang Y, Yang C, Brown JB, Antic T, Sethi I, et al. Quantitative analysis of multiparametric prostate MR images: differentiation between prostate cancer and normal tissue and correlation with Gleason score—a computer-aided diagnosis development study. *Radiology*. 2013; 267: 787–796. <https://doi.org/10.1148/radiol.13121454> PMID: 23392430
35. Moradi M, Salcudean SE, Chang SD, Jones EC, Buchan N, Casey RG, et al. Multiparametric MRI maps for detection and grading of dominant prostate tumors. *J Magn Reson Imaging JMIR*. 2012; 35: 1403–1413. <https://doi.org/10.1002/jmri.23540> PMID: 22267089
36. Mahapatra D, Buhmann JM. Prostate MRI segmentation using learned semantic knowledge and graph cuts. *IEEE Trans Biomed Eng*. 2014; 61: 756–764. <https://doi.org/10.1109/TBME.2013.2289306> PMID: 24235297
37. Kim TH, Jeong JY, Lee SW, Kim CK, Park BK, Sung HH, et al. Diffusion-weighted magnetic resonance imaging for prediction of insignificant prostate cancer in potential candidates for active surveillance. *Eur Radiol*. 2015; 25: 1786–1792. <https://doi.org/10.1007/s00330-014-3566-2> PMID: 25636416
38. Koo JH, Kim CK, Choi D, Park BK, Kwon GY, Kim B. Diffusion-weighted magnetic resonance imaging for the evaluation of prostate cancer: optimal B value at 3T. *Korean J Radiol*. 2013; 14: 61–69. <https://doi.org/10.3348/kjr.2013.14.1.61> PMID: 23323032
39. Kim CK, Park BK, Kim B. High-b-value diffusion-weighted imaging at 3 T to detect prostate cancer: comparisons between b values of 1,000 and 2,000 s/mm<sup>2</sup>. *AJR Am J Roentgenol*. 2010; 194: W33–37. <https://doi.org/10.2214/AJR.09.3004> PMID: 20028888
40. Metens T, Miranda D, Absil J, Matos C. What is the optimal b value in diffusion-weighted MR imaging to depict prostate cancer at 3T? *Eur Radiol*. 2012; 22: 703–709. <https://doi.org/10.1007/s00330-011-2298-9> PMID: 21971824
41. Roethke MC, Kuru TH, Schultze S, Tichy D, Kopp-Schneider A, Fenchel M, et al. Evaluation of the ESUR PI-RADS scoring system for multiparametric MRI of the prostate with targeted MR/TRUS fusion-guided biopsy at 3.0 Tesla. *Eur Radiol*. 2014; 24: 344–352. <https://doi.org/10.1007/s00330-013-3017-5> PMID: 24196383
42. Harnden P, Naylor B, Shelley MD, Clements H, Coles B, Mason MD. The clinical management of patients with a small volume of prostatic cancer on biopsy: what are the risks of progression? A systematic review and meta-analysis. *Cancer*. 2008; 112: 971–981. <https://doi.org/10.1002/cncr.23277> PMID: 18186496
43. Gottipati S, Warncke J, Vollmer R, Humphrey PA. Usual and unusual histologic patterns of high Gleason score 8 to 10 adenocarcinoma of the prostate in needle biopsy tissue. *Am J Surg Pathol*. 2012; 36: 900–907. <https://doi.org/10.1097/PAS.0b013e3182495dee> PMID: 22367295



44. Sadimin ET, Khani F, Diolombi M, Meliti A, Epstein JI. Interobserver Reproducibility of Percent Gleason Pattern 4 in Prostatic Adenocarcinoma on Prostate Biopsies. *Am J Surg Pathol*. 2016; 40: 1686–1692. <https://doi.org/10.1097/PAS.0000000000000714> PMID: 27487740
45. El-Shater Bosaily A, Valerio M, Hu Y, Freeman A, Jameson C, Brown L, et al. The concordance between the volume hotspot and the grade hotspot: a 3-D reconstructive model using the pathology outputs from the PROMIS trial. *Prostate Cancer Prostatic Dis*. 2016; 19: 258–263. <https://doi.org/10.1038/pcan.2016.7> PMID: 27401032
46. Winters BR, Wright JL, Holt SK, Lin DW, Ellis WJ, Dalkin BL, et al. Extreme Gleason Upgrading From Biopsy to Radical Prostatectomy: A Population-based Analysis. *Urology*. 2016; 96: 148–155. <https://doi.org/10.1016/j.urology.2016.04.058> PMID: 27313123
47. Zar JH. *Biostatistical Analysis*. 5 edition. Upper Saddle River, New Jersey, USA: Pearson; 2010.
48. Rabah DM, Arafa MA. Prostate cancer screening in a Saudi population: an explanatory trial study. *Prostate Cancer Prostatic Dis*. 2010; 13: 191–194. <https://doi.org/10.1038/pcan.2009.60> PMID: 20066007
49. Hugosson J, Aus G, Bergdahl S, Fernlund P, Frösing R, Loddning P, et al. Population-based screening for prostate cancer by measuring free and total serum prostate-specific antigen in Sweden. *BJU Int*. 2003; 92 Suppl 2: 39–43.
50. Kasel-Seibert M, Lehmann T, Aschenbach R, Guettler FV, Abubrig M, Grimm M-O, et al. Assessment of PI-RADS v2 for the Detection of Prostate Cancer. *Eur J Radiol*. 2016; 85: 726–731. <https://doi.org/10.1016/j.ejrad.2016.01.011> PMID: 26971415
51. Giannini V, Mazzetti S, Armando E, Carabalona S, Russo F, Giacobbe A, et al. Multiparametric magnetic resonance imaging of the prostate with computer-aided detection: experienced observer performance study. *Eur Radiol*. 2017; <https://doi.org/10.1007/s00330-017-4805-0> PMID: 28386721
52. Litjens GJS, Barentsz JO, Karssemeijer N, Huisman HJ. Clinical evaluation of a computer-aided diagnosis system for determining cancer aggressiveness in prostate MRI. *Eur Radiol*. 2015; 25: 3187–3199. <https://doi.org/10.1007/s00330-015-3743-y> PMID: 26060063

## 5 Diskussion

In dieser Studie wurde die wachsende Notwendigkeit für kommerzielle CAD-Systeme für den Bedarf im radiologischen und gegebenenfalls auch im urologischen Alltag gezeigt. Eine nicht invasive automatisierte Vorhersageoption des Risikoprofils einer Prostataläsion wäre im diagnostischen Alltag sehr wünschenswert. Es gibt immer mehr Studien, welche die Rolle von CAD-Systemen in Bezug auf die Unterstützung von radiologischen Befunden untersuchen. Groß angelegte Studien ( $n = 89$  (Giannini et al. 2017) und  $n = 107$  (Litjens et al. 2015)) zeigen unterschiedliche Wege, wie CAD-Systeme die Befundung der Radiologen verbessern könnten. Die Sensitivität hierbei konnte von 80,9% auf 87,6% verbessert werden (Giannini et al. 2017).

Unter Zuhilfenahme von CAD-Systemen kann die Differenzierung sowohl zwischen benignen und malignen (AUC 0,81 und 0,88), als auch zwischen indolenten und aggressiven (AUC 0,78 gegen 0,88) Tumoren verbessert werden (Litjens et al. 2015). In beiden Originalarbeiten wurde das CAD-System jedoch zuvor mit vergleichbaren Datensätzen aufgebaut und abgestimmt. In unserer Studie war dies aufgrund der fehlenden Möglichkeit der Interaktion mit den hinterlegten CAD-Informationen nicht möglich. Die diagnostische Genauigkeit der Watson Elementary™ Software wurde in einem Kollektiv von 79 Patienten mit insgesamt 47 malignen und 57 benignen MRT-gestützt biopsierten Prostataläsionen evaluiert.

Watson Elementary™ bearbeitet automatisch die Bildserien des multiparametrischen Prostata-MRT und errechnet einen Malignitätsindex (MAI), welcher die Malignomwahrscheinlichkeit einer Läsion vorhersagen kann. In der Vorhersage eines Prostatakarzinoms erreichte die Software in unserer Studie ohne Prädilektionskriterien oder Cut-off-Werte eine Sensitivität von 46,81%, eine Spezifität von 75,44% und einen positiv prädiktiven Wert von 61,11%. In einer vorherigen Studie von Roethke et al. 2016 wurde die diagnostische Genauigkeit von Watson Elementary™ an 45 Patienten mit insgesamt 1102 MR/TRUS-biopsierten Läsionen (76 maligne, 1026 benigne) untersucht. Hier erreichte die Software mit einem bestmöglichen Cut-off-Wert des quantitativen Malignitätsindex eine Sensitivität von 85,71% sowie eine Spezifität von 87,50%. Nachdem bei unserer Studie ein optimierter Cut-off-Wert verwendet wurde, konnte eine Sensitivität/Spezifität von 70,2%/61,4% erreicht werden. Diese Ergebnisse liegen noch immer unter den publizierten Daten der Studie von Roethke et al. 2016, aber sie sind dennoch bezüglich der Methoden besser vergleichbar. Die Diver-



genz zwischen diesen beiden Studien liegt an mehreren, sich unterscheidenden Einflüssen.

Ein Faktor stellt die unterschiedliche Probengewinnung dar. So wurde bei Roethke et al. 2016 eine Kombination aus MRT-gestützten transperinealen und systemischen Biopsien (ca. 25 Biopsien pro Patient) verwendet. In unserer Arbeit wurden die Proben ausschließlich durch MRT-gestützte transrektale Biopsien (ca. zwei Proben pro Patient) gewonnen. Kapitel 2.5 beschreibt beide Methoden. Hieraus wird ersichtlich, dass die MRT-gestützte Biopsie die genauere Biopsieform darstellt. Dadurch ist eine genauere Lokalisation der Pathologie möglich, wohingegen durch die systemische Biopsie keine genaue Aussage zur Lokalisation der Histologie erfolgen kann. Abgesehen von der Anzahl der entnommenen Proben, definieren beide Arbeiten eine Läsion als „maligne“, wenn sich mindestens ein Untersuchungsstück als positiv für Malignität bestätigt. Ein weiterer Unterschied in der Analyse, welcher die Abweichung der Ergebnisse verursachen kann, findet sich in der ausgewählten Schwelle des Malignitätsindex (MAI-Wert) zur Definition einer malignen Läsion. Roethke et al. 2016 legte diesen Wert bei 0,6 fest, denn dieser sollte einem Gleason Score von 6 entsprechen. Die Ergebnisse unserer Studie zeigen jedoch, dass die MAI-Werte bei fast allen Läsionen, unabhängig vom Charakter, größer als 0,6 sind. Eine Voraussetzung für eine hohe CAD-Genauigkeit, insbesondere der Neuron Network basierten Softwaren wie Watson Elementary™, ist das vielfältige Training des Systems in mehreren Ebenen und an möglichst vielen zu untersuchenden Datensätzen mit ähnlichen Ergebnissen (Schmidhuber. 2015., Song et al. 2017).

Ein großer Nachteil und eine Begrenzung für unsere Studie war die fehlende Möglichkeit der Anpassung beziehungsweise des Trainings des hinterlegten CAD-Algorithmus an unseren Datensatz. Roethke et al. 2016 haben das System grundsätzlich etabliert und mit ähnlichen Daten, entsprechend ihrem Testkollektiv im Voraus trainiert (3T Siemens Scanner). Mit diesem Training des Neuron Networks vor dem Testlauf des Programms sind somit zum Teil die Leistungsunterschiede zu unserer Studie zurückzuführen. So wurde in unterschiedlichen Publikationen gezeigt, dass mit einem aktiven überwachten Lernen der Datensätze verbesserte Ergebnisse mittels Neural Network erzielt werden können (Fei et al. 2017, Song et al. 2015, Greer et al. 2018).

Parameter wie Spulen, Magnetfeldinhomogenitäten und Untersuchungsprotokolle

führen zu Kontrastdifferenzen, welche vor der Etablierung der CAD-Software an einem neuen Arbeitsplatz durch intensives Training beseitigt werden sollen. In einer vorangegangenen Studie von Wang et al. 2014 wurden verschiedene Arbeiten in Bezug auf differente Merkmale (Feldstärke, Analysemodalitäten, Definition der Läsionen und angewandte Systeme) mit bis zu 100 MRT-Fällen verglichen. Hierbei wurde ein breites Spektrum an Ungleichheiten aufgezeigt, welche die Notwendigkeit des Trainings mit unterschiedlichen Quellen hervorheben. Der Einsatz unterschiedlicher Patientenspulen, wie die Endorektalspule erhöht die Variation der Untersuchungsmethodik. Weiterhin wurden in der vorherigen Studie b-Werte zwischen 0-800/mm<sup>2</sup> verwendet (Roethke M C et al. 2016), wohingegen in unserer Studie b-Werte von 0-1000 s/mm<sup>2</sup> genutzt wurden. Manenti et al. 2014 zeigte eine Verbesserung der Genauigkeit bei b-Werten von 1000-2000 s/mm<sup>2</sup>, sodass die Verwendung von b-Werten bis lediglich 800 s/mm<sup>2</sup> kritisch hinterfragt werden sollte. Diese Differenz kann die ADC-Wert-Berechnung deutlich beeinflussen und einen signifikanten Unterschied in der Auswertung erbringen.

Das Volumen der Läsionen sollte gesondert betrachtet werden. Die vorliegende Studie hat als ein Ausschlusskriterium eine technische Unstimmigkeit mit der CAD-Software. So lagen in circa 86% der Fälle Läsionen unter 1 ml vor, in 49% der Fälle sogar kleiner als 0,5ml. Watson Elementary™ zeigte lediglich eine zufriedenstellende Vorhersageleistung bei Läsionen größer als 1,0 ml. Dies erklärt gegebenenfalls die variierenden Ergebnisse der Studien, bei denen die Untersuchungsvolumina unterschiedlich waren. In der Vergleichsstudie von Roethke et al. 2016 wurden Läsionen unter 0,5 ml, unter Beachtung der ESUR Richtlinien (Barentsz et al. 2012) von 2012 als „klinisch nicht signifikante Malignome“ klassifiziert und deshalb von der Studie ausgeschlossen. Hervorzuheben ist jedoch, dass maligne Läsionen kleiner als 0,5 ml keinen Unterschied in Bezug auf Malignität und den Gleason Score im Vergleich zu größeren Läsionen zeigen. (Thon et al. 2017, Figur 7c) Diese Ergebnisse unterstützen die Evidenzlage, dass auch kleinere Läsionen einen signifikanten Einfluss auf das Outcome und den Behandlungserfolg der Patienten haben können. Dies sollte dazu beitragen, auch kleinere Defekte gemäß den Leitlinien zu biopsieren und eventuell zu therapieren (Harnden et al. 2008).

Des Weiteren stellt die große Variation der Gesamtheit der Prostatapathologien einen wichtigen Einflussfaktor, welcher die Ergebnisse beeinflusst hat, dar. Das CAD-

System Watson Elementary™ wurde dahingehend entwickelt, dass eine Übereinstimmung mit dem Gleason Score errechnet werden soll (Roethke et al. 2016). Dies ist durch unsere Studie nicht bestätigt worden. Auch hier können Variationen zwischen den einzelnen auszuwertenden Datensätzen (Training und Testung) schon zu Unterschieden in der Bewertung des Gleason Scores führen. Wie vorangegangene Studien zeigen, besteht eine minimale Abweichung zwischen unterschiedlichen Begutachtern meist nur im oberen oder unteren Bereich (4 und 8 bis 10) des Gleason Scores (Gottipati et al. 2012, Sadimin et al. 2016).

Außerdem spielt die Verteilung des Gleason Scores der untersuchten Läsionen eine weitere wichtige Rolle bezüglich der statistischen Auswertungen. So lagen in unserer Studie ein Großteil der Gleasonwerte zwischen 6 oder 7 (3+3 oder 3+4) mit weniger Läsionen im niedrigen und im oberen Scala-Bereich. Diese ungleichmäßige Verteilung beeinflusst die Pearson Korrelation möglicherweise negativ, so dass eine Korrelation zwischen Malignitätsindex und Gleason Score statistisch nicht nachgewiesen werden kann (Zar. 2010).

Die ROC-Analyse der PI-RADS™ Klassifikation, zeigt eine geringere Genauigkeit in unserer Studie im Vergleich zu vorher publizierten Arbeiten auf (Bittencourt et al. 2015, Roethke et al. 2014). Eine vorangegangene Studie von Kasel-Seibert et al. verglich die Ergebnisse von PI-RADS™ Version 2 mit Version 1. Sie stellte eine hohe Genauigkeit der AUC von 0,88 und 0,91 für V1 und V2 dar. Dennoch dürfen erhebliche Unterschiede in der Methodik und Datenauswahl keinesfalls außer Acht gelassen werden. Auf der einen Seite wurden Patienten mit PI-RADS 1 oder 2 Läsionen nicht biopsiert und somit aus unserer Studie ausgelassen, welches die Anzahl der richtig negativen Läsionen beeinflussen könnte. Der Großteil der PI-RADS 3 Läsionen wurde ebenso außen vor gelassen, da diese auch nicht biopsiert wurden. Eine geringe Anzahl der beachteten Läsionen (auch PI-RADS 2 und 3) resultiert hauptsächlich aus der Herunterstufung der PI-RADS 4 Läsionen nachdem die V2 eingeführt wurde. Auf der anderen Seite wurden in anderen Studien PI-RADS 3 Läsionen in der offiziellen Auswahl berücksichtigt (Litjens et al. 2014). Wichtig hierbei zu beachten ist, dass die Malignitätsrate mit nur 29% in den untersuchten Läsionen relativ niedrig ausfällt (Kasel-Seibert et al. 2014). In unserer Studie lag der Anteil invasiver Karzinome mit einer Malignitätsrate von 45% der biopsierten Läsionen deutlich höher.

## 6 Schlussfolgerung

Die vorliegende Studie zeigt, dass eine sorgfältig programmierte kommerzialisierte CAD-Software Watson Elementary™ bei Veränderungen der Ausstattung und des Untersuchungsprotokolls nicht zufriedenstellend operiert, obwohl fast doppelt so viele Patienten wie in vorangegangenen Studien untersucht wurden (Roethke et al. 2016, Peng et al. 2013, Shah et al. 2012).

Die fehlende Möglichkeit des Programmtrainings, die geringe Anzahl von PI-RADS 2 und 3 Läsionen und der anspruchsvolle Charakter des Datensatzes, einschließlich kleiner Läsionen, welche in anderen publizierten Studien als Ausschlusskriterium festgelegt wurden, sollten als mögliche beeinflussende Faktoren betrachtet werden. Wichtig zu beachten ist jedoch, dass die mpMRT-Parameter, welche in unserer Abteilung angewendet werden, mit den Empfehlungen des American College of Radiology in PI-RADS™ V2 übereinstimmen. Im Vergleich zu einer vorhergehenden Studie von Wang et al. unterstützen unsere Ergebnisse die Annahme, dass die diagnostische Bedeutung des getesteten CAD-Systems stark vom untersuchten Datensatz abhängig ist (Wang et al. 2014).

In Zusammenschau aller Ergebnisse ist die getestete CAD-Software nicht zufriedenstellend für die alleinige Nutzung. Eine notwendige und wichtige Voraussetzung für eine CAD-Software ist die Zuverlässigkeit der diagnostischen Aussagefähigkeit vom Datensatz der unterschiedlichen Ausgangsdaten. Eine breit gefächerte Grundlage an Datensätzen für besseres und eventuell auch interaktives Systemtraining sollte zukünftig aktiver in die Software integriert werden.

## 7 Literatur- und Quellenverzeichnis

Amann K, Kain R, Klöppel G. 2016. Pathologie: Urogenitale Und Endokrine Organe, Gelenke Und Skelett. 3. Aufl. Berlin, Heidelberg: Springer-Verlag, 140, 142, 149-151, 155-157, 169, 171, 177-178.

Attenberger U, Ritter M, Wenz F. 2017. MR- und PET-Bildgebung der Prostata Diagnostik und Therapieplanung. 1. Auflage. Berlin, Heidelberg: Springer-Verlag, 4-10.

Barentsz J O, Richenberg J, Clements R, Choyke P, Verma G, Villeirs G, Rouviere O, Logager V, Fütterer J J. 2012. ESUR prostate MR guidelines 2012. European Radiology, 22 (4): 746-757. DOI 10.1007/s00330-011-2377-y.

Barentsz J O, Weinreb J C, Verma S, Thoeny H C, Tempany C M, Shtern F, Padhani A R, Margolis D, Macura K J, Haider M A, Cornud F, Choyke P L. 2016. Synopsis of the PI-RADS v2 Guidelines for Multiparametric Prostate Magnetic Resonance Imaging and Recommendations for Use. European Urology 6 (1): 41-49. DOI 10.1016/j.eururo.2015.08.038.

Barnes B, Bertz J, Buttman-Schweiger N, Fiebig J, Jordan S et al. 2016. Bericht zum Krebsgeschehen in Deutschland 2016. RKI, 22-23, 40-43.

Barrett T, Turkbey B, Choyke P L. 2015. PI-RADS version 2: what you need to know. Clinical Radiology 7 (11): 1165-1176. DOI 10.1016/j.crad.2015.06.093.

Bittencourt L K, Litjens G, Hulsbergen-van de Kaa C A, Turkbey B, Gasparetto E L, Barentsz J O. 2015. Prostate Cancer: The European Society of Urogenital Radiology Prostate Imaging Reporting and Data System Criteria for Predicting Extraprostatic Extension by Using 3-T Multiparametric MR Imaging. Radiology 276 (2): 479-489. DOI 10.1148/radiol.15141412.

Bjurlin M A, Meng X, Le Nobin J, Wysock J S, Lepor H, Rosenkrantz A B, Taneja S. 2014. Optimization of Prostate Biopsy: the Role of Magnetic Resonance Imaging Targeted Biopsy in Detection, Localization and Risk Assessment. The Journal of Urology 192 (3): 648-658. DOI 10.1016/j.juro.2014.03.117

Börgermann C, Chun F K-H, Bastian P J, Fischer K, Fornara P, Fröhner M, Graefen M, Haese A, Hammerer P, Heine K, Heß J, Hoda M R, Huland H, Köllermann J, Loertzer H, Luboldt H J, Miller K, Rübber H, Schlomm T, Schostak M, Schrader M,

- Schwarz R, Thederan I, Wirth M, Wolff J M, Hansen J. 2014. Prostatakarzinom. Rübben H. (eds) Uroonkologie. Berlin, Heidelberg: Springer-Verlag, 513-676.
- Brierley J D, Gospodarowicz M K, Wittekind C. 2017. TNM Classification of Malignant Tumours. 8. Auflage. Union for International Cancer Control. Oxford: Wiley Blackwell.
- Brock M, von Bodmann C, Palisaar J, Becker W, Martin-Seidel P, Noldus J. 2015. Detecting prostate cancer—a prospective comparison of systematic prostate biopsy with targeted biopsy guided by fused MRI and transrectal ultrasound. *Deutsches Ärzteblatt International* 112 (37): 605-611. DOI 10.3238/arztebl.2015.0605.
- Fei Y, Hu J, Gao K, Tu J, Li W Q, Wang W. 2017. Predicting risk for portal vein thrombosis in acute pancreatitis patients: A comparison of radical basis function artificial neural network and logistic regression models. *Journal of Critical Care* 39 (Juni 2017): 115-123. DOI 10.1016/j.jcrc.2017.02.032.
- Gheewala T, Troy S, Gnanasekar M. 2017. Photosensitizers in Prostate Cancer Therapy. *Oncotarget* 8 (18): 30524–30538. DOI 10.18632/oncotarget.15496
- Giannini V, Mazzetti S, Armando E, Carabalona S, Russo F, Giacobbe A, Muto G, Regge D. 2017. Multiparametric magnetic, resonance imaging of the prostate with computer-aided detection: experienced observer performance study. *European Radiology* 27 (10): 4200-4208. DOI 10.1007/s00330-017-4805-0.
- Gottipati S, Warncke J, Vollmer R, Humphrey P A. 2012. Usual and unusual histologic patterns of high Gleason score 8 to 10 adenocarcinoma of the prostate in needle biopsy tissue. *The American Journal of Surgical Pathology* 36 (6): 900-907. DOI 10.1097/PAS.0b013e3182495dee.
- Greer H, Gerber S, Niethammer M, Kwitt R, McCormick M, Chittajallu D, Siekierski N, Oetgen M, Cleary K, Aylward S. 2018. Scoliosis screening and monitoring using self contained ultrasound and neural networks. *IEEE International Symposium on Biomedical Imaging* (April 2018): 1500-1503. DOI 10.1109/ISBI.2018.8363857.
- Harnden P, Naylor B, Shelley S D, Clements H, Coles B, Mason M D. 2008. The clinical management of patients with a small volume of prostatic cancer on biopsy: what are the risks of progression? A systematic review and meta-analysis. *Cancer* 112 (5): 971-981. DOI 10.1002/cncr.23277.

- Hegde J V, Mulkern R V, Panych L P, Fennessy F M, Fedorov A, Maier S E, Tempany C M C. 2013. Multiparametric MRI of Prostate Cancer: An Update on State-of-the-Art Techniques and Their Performance in Detecting and Localizing Prostate Cancer. *Journal of Magnetic Resonance Imaging* 37 (5): 1035-1054. DOI 10.1002/jmri.23860.
- Heidenreich A, Bastian P J, Bellmunt J, Bolla M, Joniau S, van der Kwast T, Mason M, Matveev V, Wiegel T, Zattoni F, Mottet N. 2014. EAU Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent—Update 2013. *European Urology*, Volume 65 (1): 124-137. DOI 10.1016/j.eururo.2013.09.046.
- Ilic D, Neuberger M M, Djulbegovic M, Dahm P. 2013. Screening for prostate cancer. *Cochrane Database Systematic Reviews*. DOI 10.1002/14651858.CD004720.pub3.
- Jocham D, Ell C, Baumgartner R, Gossner L, Häußinger K, Iro H, Szeimies R-M. 2000. Photodynamische Therapie. *Deutsches Ärzteblatt* 97 (49): A3337-A3343.
- Kasel-Seibert M, Lehmann T, Aschenbach R, Guettler F V, Abubrig M, Grimm M O, Teichgraeber U, Franiel T. 2016. Assessment of PI-RADS v2 for the Detection of Prostate Cancer. *European Journal of Radiology* 85 (4): 726-731. DOI 10.1016/j.ejrad.2016.01.011.
- Kirkham A P S, Emberton M, Allen C. 2006. How good is MRI at detecting and characterising cancer within the prostate?. *European Urology* 50 (6): 1163- 1175. DOI 10.1016/j.eururo.2006.06.025.
- Klotz L. 2017. Active surveillance for low-risk prostate cancer. *Current Opinion in Urology* 27 (3): 225–230. DOI 10.1097/MOU.0000000000000393.
- Kozlowski P, Chang S D, Jones E C, Berean K W, Chen H, Goldenberg S L. 2006. Combined diffusion-weighted and dynamic contrast-enhanced MRI for prostate cancer diagnosis-Correlation with biopsy and histopathology. *Journal of Magnetic Resonance Imaging JMRI* 24: 108-113. DOI 10.1002/jmri.20626.

Leitlinienprogramm Onkologie der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V. (AWMF), Deutsche Krebsgesellschaft e.V. (DKG), Deutsche Krebshilfe (DKH). 2018. In: Interdisziplinäre Leitlinie der Qualität S3 zur Früherkennung, Diagnose und Therapie der verschiedenen Stadien des Prostatakarzinoms, Langversion 5.0.

Litjens G J S, Barentsz J O, Karssemeijer N, Huisman HJ. 2015. Clinical evaluation of a computer-aided diagnosis system for determining cancer aggressiveness in prostate MRI. *European Urology* 25 (11): 3187-3199. DOI 10.1007/s00330-015-3743-y.

Litjens G, Debats O, Barentsz J, Karssemeijer N, Huisman H. 2014. Computer-Aided Detection of Prostate Cancer in MRI. *IEEE Transactions on Medical Imaging* 33 (5): 1083-1092. DOI 10.1109/TMI.2014.2303821.

Lu-Yao G L, Albertsen P C, Moore D F, Shih W, Lin Y, DiPaola R S, Yao S-L. 2014. Fifteen-Year Survival Outcomes Following Primary Androgen-Deprivation Therapy for Localized Prostate Cancer. *JAMA Intern Med* 174 (9): 1460–1467. DOI 10.1001/jamainternmed.2014.3028.

Manenti G, Nezzo M, Chegai F, Vasili E, Bonanno E, Simonetti G. 2014. DWI of Prostate Cancer: Optimal b-Value in Clinical Practice. *Prostate Cancer* 2014. DOI 10.1155/2014/868269.

Moore C M, Ridout A, Emberton M. 2013. The role of MRI in active surveillance of prostate cancer. *Current Opinion in Urology* 23 (3): 261–267. DOI 10.1097/MOU.0b013e32835f899f.

Mottet N, Bellmunt J, Bolla M, Briers E, Cumberbatch M G, De Santis M, Fossati N, Gross T, Henry A M, Joniau S, Lam T B, Mason M D, Matveev V B, Moldovan P C, van den Bergh R C N, Van den Broeck T, van der Poel H G, van der Kwast T H, Rouvière O, Schoots I G, Wiegel T, Cornford P. 2017. EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. *European Urology* 71 (4): 618-629. DOI 10.1016/j.eururo.2016.08.003.



- Ning P, Shi D, Sonn G A, Vasanawala S S, Loening A M, Ghanouni P, Obara P, Shin L K, Fan R E, Hargreaves B A, Daniel B L. 2018. The impact of computed high b-value images on the diagnostic accuracy of DWI for prostate cancer: A receiver operating characteristics analysis. *Scientific Reports* (8). DOI 10.1038/s41598-018-21523-6
- Nörenberg D, Solyanik O, Schlenker B, Magistro G, Ertl-Wagner B, Clevert D A, Stief C, Reiser M F, D'Anastasi M. 2017. MRT der Prostata. *Der Urologe* 56 (5): 665-677. DOI 10.1007/s00120-017-0378-4.
- Peng Y, Jiang Y, Yang C, Brown J B, Antic T, Sethi I, Schmid-Tannwald C, Giger M L, Eggenner S E, Oto A. 2013. Quantitative analysis of multiparametric prostate MR images: Differentiation between Prostate Cancer and Normal Tissue and Correlation with Gleason Score- A computer-aided Diagnosis Development Study. *Radiology* 267 (3): 787-796. DOI 10.1148/radiol.13121454.
- Ritch C R. 2017. Photodynamic therapy for low risk prostate cancer. *BMJ* 356 (575). DOI 10.1136/bmj.j575
- Roethke M C, Kuru T H, Mueller-Wolf M B, Agterhuis E, Edler C, Hohenfellner M, Schlemmer H-P, Hadaschik B A. 2016. Evaluation of an Automated Analysis Tool for Prostate Cancer Prediction Using Multiparametric Magnetic Resonance Imaging. *PLoS ONE* 11 (7): e0159803. DOI 10.1371/journal.pone.0159803.
- Roethke M C, Kuru T H, Schultze S, Tichy D, Kopp-Schneider A, Fenchel M, Schlemmer H-P, Hadaschik B A. 2014. Evaluation of the ESUR PI-RADS scoring system for multiparametric MRI of the prostate with targeted MR/TRUS fusion-guided biopsy at 3.0 Tesla. *European Radiology* 24 (2): 344-352. DOI 10.1007/s00330-013-3017-5.
- Sadimin E T, Khani F, Diolombi M, Meliti A, Epstein J I. 2016. Interobserver Reproducibility of Percent Gleason Pattern 4 in Prostatic Adenocarcinoma on Prostate Biopsies. *The American Journal of Surgical Pathology* 40 (12): 1686-1692. DOI 10.1097/PAS.0000000000000714.
- Schmidhuber J. 2015. Deep learning in neural networks: An overview. *Neural networks* 61: 85-117. DOI 10.1016/j.neunet.2014.09.003.

- Shah V, Turkbey B, Mani H, Pang Y, Pohida T, Merino M J, Pinto P A, Choyke P L, Bernardo M. 2012. Decision support system for localizing prostate cancer based on multiparametric magnetic resonance imaging. *Medical Physics* 39 (7): 4093-4103. DOI 10.1118/1.4722753.
- Song H, Thiagarajan J J, Sattigeri P, Spanias A. 2017. Optimizing Kernel Machines using Deep Learning. *IEEE transaction on neural networks and learning Systems*. DOI 1711.05374v1.
- Song W, Li S, Liu J, Qin H, Zhang B, Shuyang Z, Hao A. 2015. Multi-task Cascade Convolution Neural Networks for Automatic Thyroid Nodule Detection and Recognition. *IEEE Journal of Biomedical and Health Informatics* 14(8). DOI 10.1109/JBHI.2018.2852718.
- Thon A, Teichgräber U, Tennstedt-Schenk C, Hadjidemetriou S, Winzler S, Malich A, Papageorgiou I. 2017. Computer aided detection in prostate cancer diagnostics: A promising alternative to biopsy? A retrospective study from 104 lesions with histological ground truth. *PLoS ONE* 12 (10). DOI 10.1371/journal.pone.0185995.
- Turbey B, Shah VP, Pang Y, Bernardo M, Xu S, Kruecker J, Locklin J, Baccala A A, Rastinehad A R, Merino M J, Shih J H, Wood B J, Pinto P A, Choyke P L. 2011. Is apparent diffusion coefficient associated with clinical risk scores for prostate cancers that are visible on 3-T MR images?. *Radiology* 258 (2): 488-495. DOI 10.1148/radiol.10100667.
- Wang S, Burt K, Turkbey B, Choyke P, Summers R M. 2014. Computer Aided-Diagnosis of Prostate Cancer on Multiparametric MRI: A Technical Review of Current Research. *BioMed Research International* 2014. DOI 10.1155/2014/789561.
- Weinreb J C, Barentsz J O, Choyke P L, Cornud F, Haider M A, Macura K J, Margolis D, Schnall M D, Shtern F, Tempany C M, Thoeny H C, Verma S. 2015. PI-RADS Prostate Imaging – Reporting and Data System: 2015, Version 2. *European Urology* 69 (1): 16-40. DOI 10.1016/j.eururo.2015.08.052.
- Wittekind C, Sobin L H, Klimpfinger M. 2005. *TNM-Atlas*. 5. Auflage. Berlin, Heidelberg: Springer-Verlag, 283-291.

Wolf A M D, Wender R C, Etzioni R B, Thompson I M, D'Amico A V, Volk R J, Brooks D D, Dash C, Guessous I, Andrews K, DeSantis C, Smith R A. 2010. American Cancer Society Guideline for the Early Detection of Prostate Cancer: Update 2010. *Cancer Journal for Clinicians* 60 (2): 70-98. DOI [org/10.3322/caac.20066](https://doi.org/10.3322/caac.20066)

Zar J H. 2010. *Biostatistical Analysis*. 5th Edition. Northern Illinois University. Pearson.

## Anhang

### Abbildungsverzeichnis

Abb. 1	Aufbau der Prostata und Zoneneinteilung nach PI-RADS V2.....	9
Abb. 2	Übersicht über die Zusammensetzung der Bewertungen in PI-RADS V2 der unterschiedlichen Zonen und Sequenzen.....	17
Abb. 3	Prostatastanzbiopsie.....	18
Abb. 4	Gleason Score.....	19
Abb. 5	Malignancy Attention Index.....	21
Abb. 6	MAI-Säulendiagramm.....	21

## **Danksagung**

Abschließend möchte ich mich bei allen Personen bedanken, die mir bei der Erstellung und Erarbeitung meiner Arbeit unterstützend zur Seite standen.

An erster Stelle geht mein Dank an meine Betreuer der Dissertation, Herrn Prof. Dr. med. habil. A. Malich, der für die Vermittlung des Themas, die Bereitstellung der Daten und die Organisation in Zusammenarbeit mit der Universität Jena, sowie mit seiner allseits bestehenden Unterstützung während der Bearbeitung jederzeit für mich da war.

An Herrn Prof. Dr. med. U. Teichgräber als Betreuer der Dissertation an der Universität Jena, der bei Fragen hilfsbereit zur Seite stand und die Bearbeitung des Themas ermöglichte.

Desweiteren gilt mein besonderer Dank Frau Dr. I. E. Papageorgiou, M. Sc., welche mir und dem Fortschreiten der Arbeit viel Zeit opferte. Durch Ihr Fachwissen in der Radiologie sowie umfassende Erfahrungen mit Publikationen und statistischen Auswertungen zeigte sie mir stets die richtige Richtung und gab mir wertvolle Tipps. Sie war mir ein sehr wichtiger Wegbegleiter.

Ein weiterer besonderer Dank gilt meiner Familie und meinem Mann für ihr entgegengebrachtes Vertrauen und ihre Unterstützung in dieser Zeit.

## **Ehrenwörtliche Erklärung**

Hiermit erkläre ich, dass mir die Promotionsordnung der Medizinischen Fakultät der Friedrich-Schiller-Universität bekannt ist,

ich die Dissertation selbst angefertigt habe und alle von mir benutzten Hilfsmittel, persönlichen Mitteilungen und Quellen in meiner Arbeit angegeben sind,

mich folgende Personen bei der Auswahl und Auswertung sowie bei der Herstellung des Manuskripts unterstützt haben: Herr Professor Dr. med. U. Teichgräber, Herr Professor Dr. Med. habil. A. Malich und Frau Dr. M. Sc. I. E. Papageorgiou,

die Hilfe eines Promotionsberaters nicht in Anspruch genommen wurde und dass Dritte weder unmittelbar noch mittelbare geldwerte Leistungen von mir für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorliegenden Dissertation stehen,

dass ich die Dissertation noch nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht habe und

dass ich die gleiche, eine in wesentlichen Teilen ähnliche oder eine andere Abhandlung nicht bei einer anderen Hochschule als Dissertation eingereicht habe.

Nordhausen, 03.09.2019

Anika Brosch